3/19

Principes van computer-elektronica

Inhoud

3/19.14 Waterkoelingstechnologie voor snelle processoren
(verschenen in de 102e aanvulling)

3/19.15 De Agent-technologie van Microsoft
(verschenen in de 106e en 107e aanvulling)

3/19.1

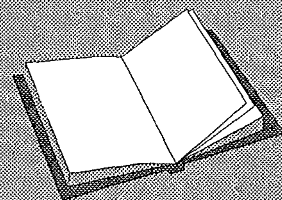
Architectuur van de computer hardware en de bus

Inleiding

"Personal" is héél letterlijk op te nemen!
Alleen al uit de naam "PC", "*Personal* Computer", kan men afleiden dat de PC géén standaard apparaat is, maar een op maat gemaakt systeem, aangepast aan de wensen van de individuele gebruiker/ster. Dat is een rechtstreeks gevolg van het feit dat de IBM-ontwerpers, die de allereerste PC op de tekentafel vorm gaven, een heel beperkt budget hadden, met veel te weinig waren en veel te weinig tijd hadden. Vandaar dat werd besloten om géén grote print te ontwerpen, waarop alle onderdelen van het ontwerp zouden zitten (standaard IBM-gebruik uit die tijd), maar een soort module-systeem. Deelschakelingen zouden onafhankelijk van elkaar ontworpen worden, op kleine printen ondergebracht worden en via een soort van connectorsysteem met een moeder-print verbonden worden. Dit uit zuinigheid en tijdgebrek geboren concept staat ongetwijfeld aan de basis van het succes van de PC! Dank zij dit connector-systeem, "bus" genoemd, en dank zij de losse printen, "kaarten" genoemd, was en is het nog steeds voor iedereen mogelijk een goedkope "kale" PC op een heel eenvoudige manier uit te breiden tot een super-de-luxe apparaat. Het volstaat goedkope kaarten uit de bus te verwijderen en deze te vervangen door veel duurdere uitvoeringen die veel meer mogelijkheden hebben. Dit "Personal"-aspect van het PC-ontwerp heeft een hele industrie in het leven geroepen van bedrijven en bedrijfjes, die alle mogelijke uitbreidingskaarten voor de PC op de markt brengen.

Systeemoverzicht
Het bespreken van de architectuur van een PC kan in feite niet los gezien worden van de toegepaste bus. Wél kan men natuurlijk wat algemeens opmerken over het minimale systeem, waaruit een werkbare hedendaagse PC moet bestaan. Een dergelijke minimale configuratie is getekend in figuur 3/19.1-1.
De minimale configuratie bestaat uit:
- de eigenlijke computer, PC genoemd;
- een magnetische eenheid voor het in- en uitlezen van gegevens;
- een toetsenbord;
- een monitor;

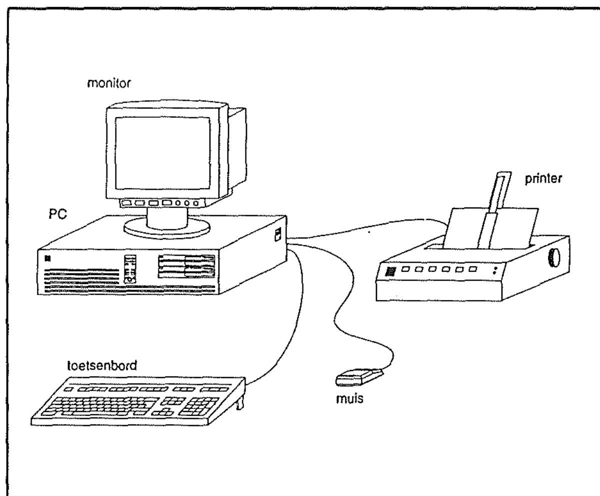
LEES OOK:
Hoofdstuk 6/9.20.1
Hoofdstuk 6/10.20



19.1 Architectuur van de computer hardware en de bus

- een printer;
- een muis.

De vijf laatstgenoemde onderdelen van het systeem, die niet in de kast van de PC hoeven te zitten, zijn ook zuiver elektronisch gezien absoluut eigen apparaten, die volledig los staan van de elektronica van de PC. Men spreekt dan ook van "perifere apparatuur" en de communicatie tussen de eigenlijke computer en deze perifere apparaten gebeurt via de bus met behulp van speciale protocollen. De bus zorgt voor de hardwarematige overdracht van de gegevens, het protocol bepaalt hoe die gegevens softwarematig worden overgedragen.



Figuur 3/19.1-1: De minimale configuratie van een moderne PC.

Iets meer elektronisch bekeken, kan men het minimale systeem van een PC voorstellen als getekend in figuur 3/19.1-2. Het hart van het systeem, de eigenlijke computer, bestaat in feite uit niet veel meer dan een processor uit Intel's 80xxx-serie, het interne RAM-geheugen en wat IC's die zorgen voor de timing van het systeem.

Alle overige apparatuur is via speciale ***adapters*** op dit basissysteem aangesloten. Voor een printer staat een parallelle adapter ter beschikking, voor het beeldscherm een grafische adapter, voor communicatie met andere PC's een netwerkadapter, etc.

Al die adapters communiceren met het basissysteem via de bus. Het zal dus duidelijk zijn dat die bus een uitzonderlijk belangrijk onderdeel van het gehele systeem is!

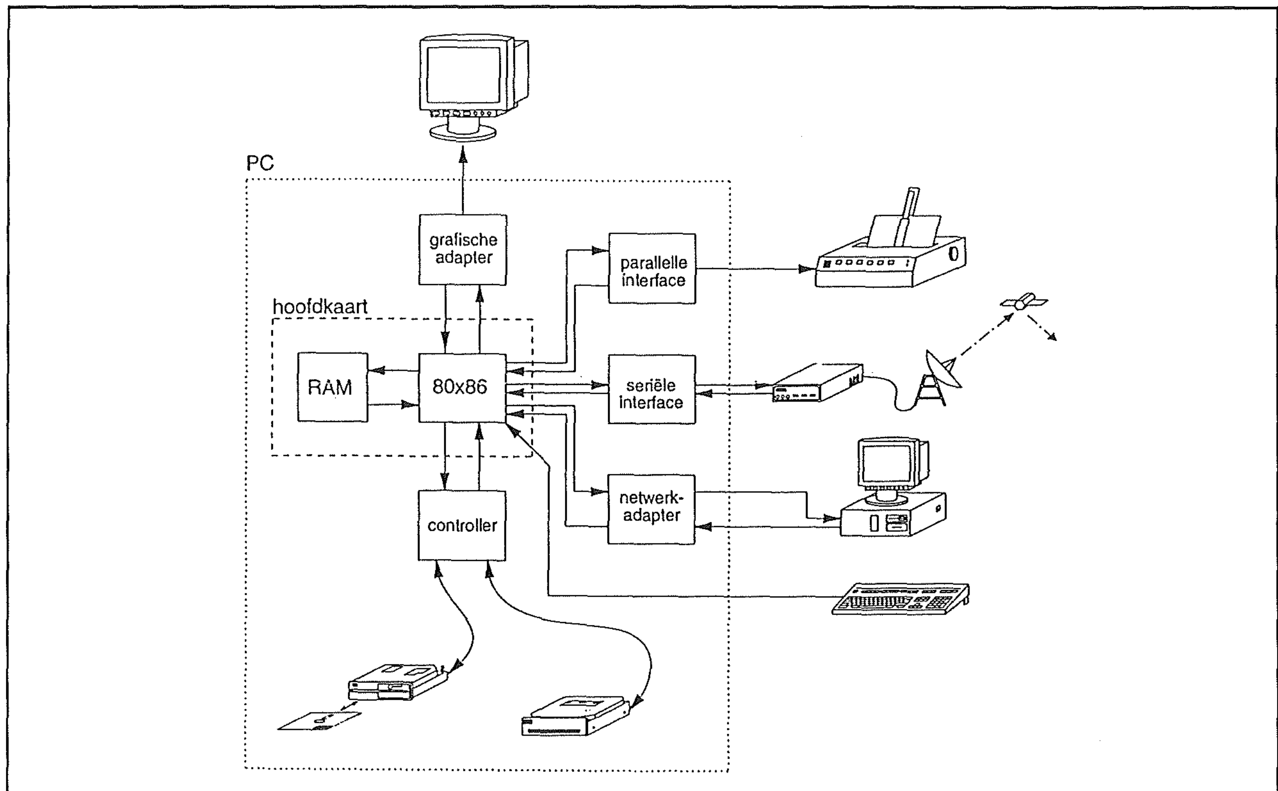
De bus

De bus vormt, zoals reeds geschreven, de verbinding tussen de afzonderlijke onderdelen van de PC. Omdat bijna alle componenten continu met elkaar moeten kunnen communiceren, is de bus van bijzonder groot belang voor de prestaties van de PC. Dank zij de structuur van de bus kan een PC naar eigen behoefte worden uitgebreid, door simpelweg uitbreidingskaarten in de daarvoor bestemde insteekplaatsen (binnen de busstructuur) te plaatsen. De enige voorwaarde is dat de uitbreidingskaart zich aan de regels van het bussysteem moet houden. Voor de rest kan de uitbreidingskaart elke gewenste taak vervullen (zoals meetgegevens verzamelen of streepjescode lezen).

Maatgevend voor de kwalitatieve beoordeling is de "breedte" van de bus: een 16 bit computer beschikt bijvoorbeeld over een bus met 16 lijnen.

De breedte van het bussysteem wordt dus in bit aangegeven, dat wil zeggen hoeveel bits parallel, dus gelijktijdig, tussen de bouwstenen kunnen worden doorgegeven. De breedte van de bus is zodoende bepalend voor de doorvoersnelheid van gegevens binnen de PC en daarmee natuurlijk ook voor de algemene verwerkingssnelheid.

19.1 Architectuur van de computer hardware en de bus



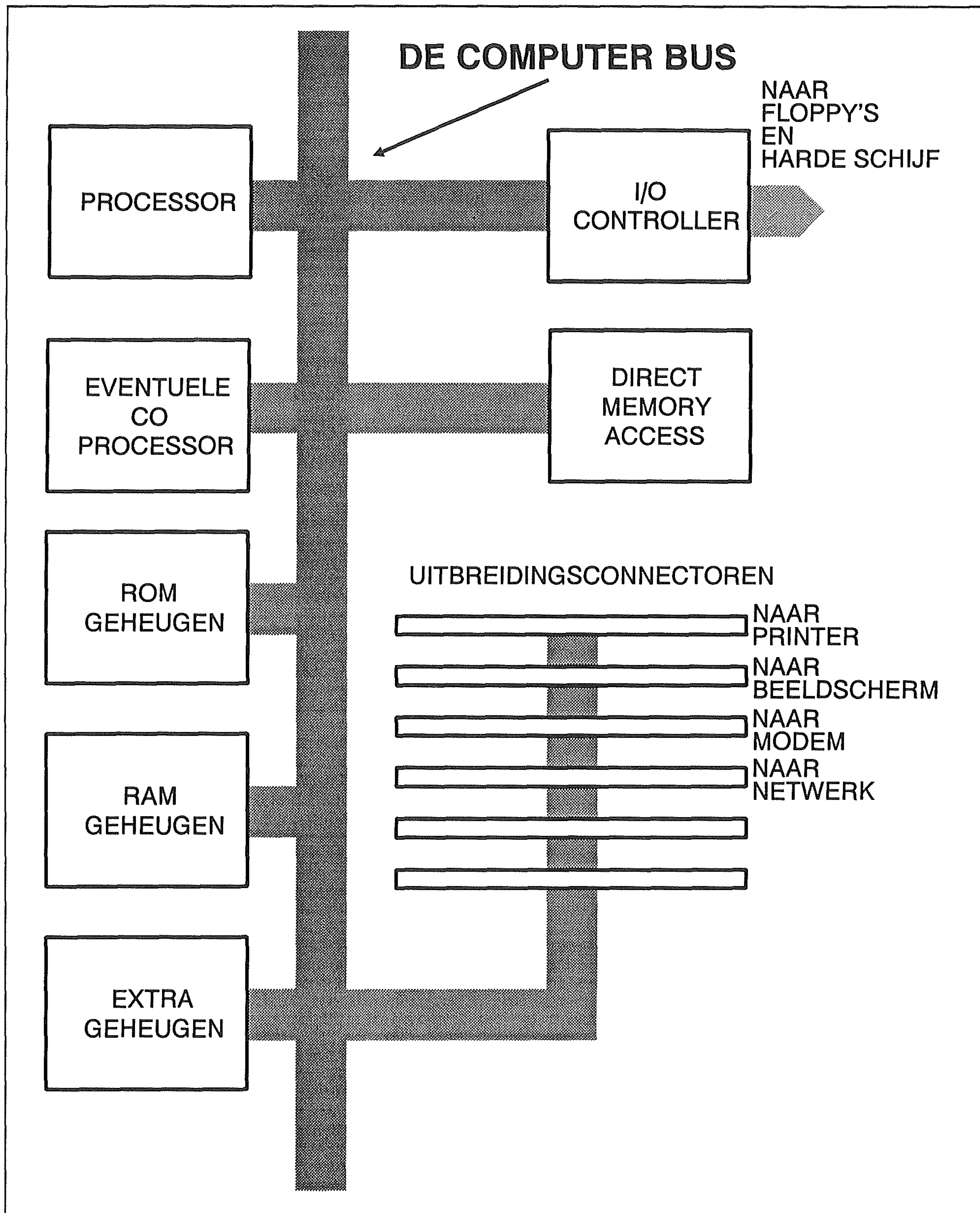
Figuur 3/19.1-2: Het systeem van een computer, met alle via de bus aangesloten adapters en overige perifere apparatuur.

De PC's van het eerste uur die met een microprocessor van het type 8088 waren uitgerust, beschikten over een 8 bit bus, 16 bit computers hebben een 80286 microprocessor en in de 32 bit machines doet een 80386 of 80486 microprocessor zijn werk. De nieuwste telg van de microprocessoren voor de PC, de Pentium, is weliswaar een 32 bit processor, maar heeft wel een 64 bit brede aansluiting op de data bus. Naast de oorspronkelijke 80386 microprocessor bestaat er ook een "afgeslankte" variant: de 386SX. De 386SX heeft slechts een 16 bit brede aansluiting op de databus, intern werkt hij (net als de "echte" 80386) wel met een 32 bit brede bus.
De schematische voorstelling van de bus binnen een computersysteem is voorgesteld in figuur 3/19.1-3.

Hieruit blijkt duidelijk dat de bus de verbindende schakel is tussen alle onderdelen van het systeem.

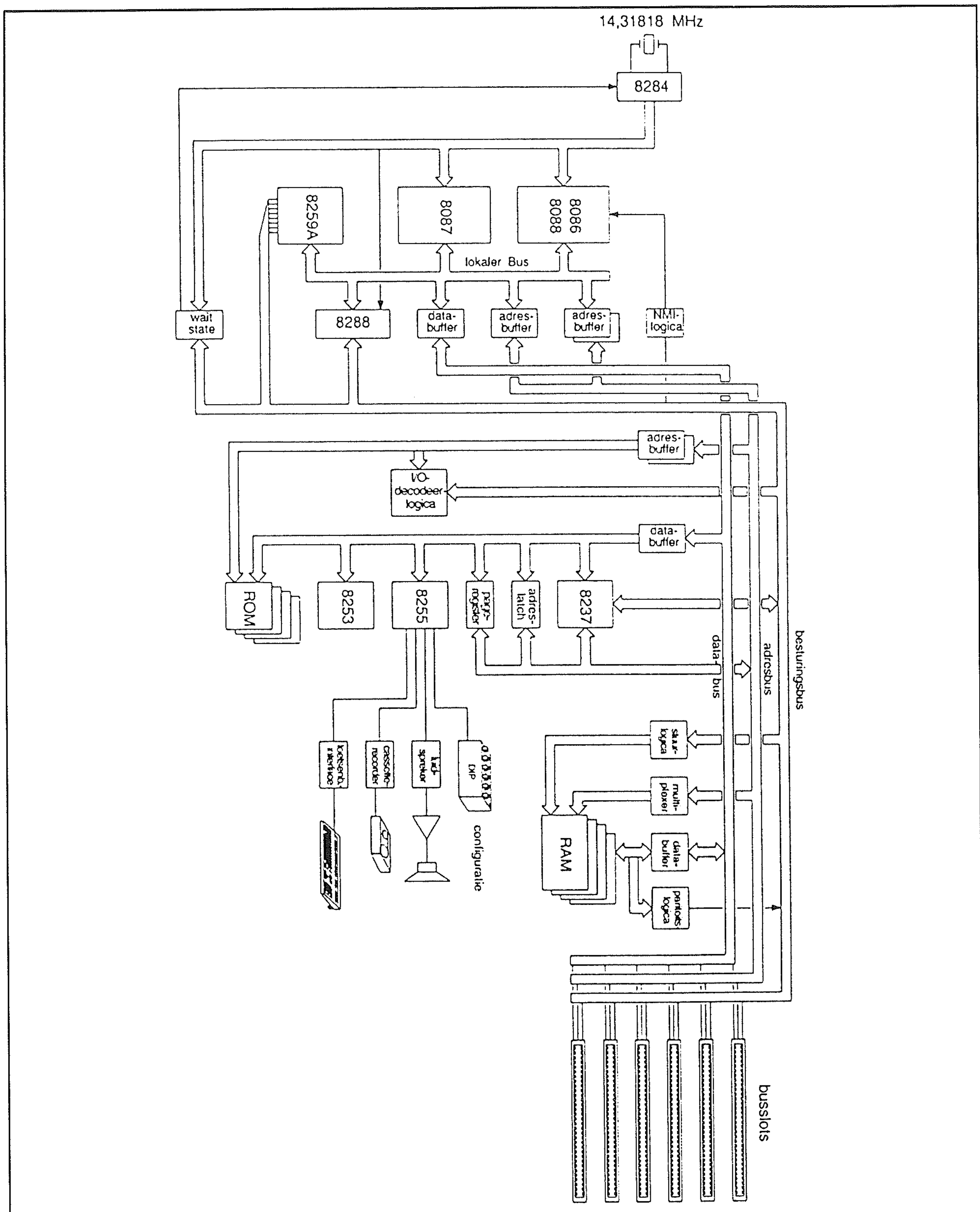
Eigenschappen van een computerbus
Naast de reeds besproken breedte van de bus wordt het systeem van de bus én van de PC bepaalt door de volgende eigenschappen.
Een bus heeft een maximale klokfrequentie die men echter niet met de klokfrequentie van de microprocessor mag verwarren. Terwijl een 80486 processor bijvoorbeeld op 66 MHz kan werken, bedraagt de klokfrequentie van een "normaal" bus-systeem slechts 8,33 MHz.
Afhankelijk van de aard van de informatie die van en naar de microprocessor wordt getransporteerd, onderscheidt men drie typen busverbindingen:

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-3: Uit deze algemene blokschematische opzet van een computer blijkt duidelijk dat de bus de verbindende schakel is tussen alle onderdelen van het systeem.

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-4: De architectuur van de PC/XT.

- De databus die de gegevens van en naar de diverse geheugens alsmede van en naar de invoer en uitvoer apparaten transporteert.
- De adresbus die de adressen waar bepaalde gegevens in het (intern) geheugen worden bewaard, de geheugenplaatsen dus, overbrengt.
- De besturingsbus (controlbus) waarover de volgorde en verloop (synchronisatie) van de processen in de centrale verwerkingseenheid worden geregeld.

Architectuur

Inleiding

De volledige architectuur van een PC-systeem wordt in hoge mate bepaald door de toegepaste bus. In de loop der tijd zijn diverse systemen ontwikkeld, die in de volgende subhoofdstukken worden besproken, namelijk:

- de PC/XT-architectuur;
- de AT- of ISA-architectuur;
- de Microchannel architectuur;
- de EISA-architectuur;
- de VESA Local Bus architectuur;
- de PCI-architectuur.

De PC/XT-architectuur

Inleiding

Deze architectuur is nu volstrekt achterhaald en een PC met deze architectuur zou niet eens geschikt zijn voor het spelen van moderne spelletjes.
Historisch gezien is deze architectuur echter van groot belang, omdat het de architectuur was waarmee IBM de eerste PC heeft uitgerust.

Blokschema van de PC/XT-architectuur

In figuur 3/19.1-4 is de blokschematische opzet van deze architectuur geschetst. Het centrale onderdeel is de processor van het type 8088 of 8086. Er is plaats voor het inbouwen van de mathematische coprocessor 8087, waardoor sommige wiskundige berekeningen sneller worden uitgevoerd, ***onder de voorwaarde dat de software de instructie-set van deze processor ondersteunt.*** De perifere schakeling 8284 wekt de systeemklok op uit een kristal met een frequentie van 14,31818 MHz. Deze frequentie wordt door drie gedeeld en levert de nu onvoorstelbaar trage systeemklok van 4,77 MHz op.
In latere uitvoeringen van de PC/XT kon, door gebruik te maken van snellere processoren, de systeemklok opgevoerd worden tot 10 MHz. Die snellere en betere processoren werden door NEC ontwikkeld en onder de typenummers V20 en V30 op de markt gebracht. Ook toen voelde Intel dus reeds de hete adem van de concurrentie in de rug!

De adresbus

De PC/XT-architectuur heeft een bus, waarvan het adres-gedeelte in vier delen gesplitst kan worden:

- De lokale adresbus bevat de 20 adressignalen van de processor. Deze signalen worden via buffers van de adresbus van het systeem gescheiden.
- De systeem adresbus is de wezenlijke adresbus voor de periferie en is van de lokale adresbus gescheiden door buffers en latches.
  Het stuursignaal om de lokale met de systeem adresbus te verbinden is het processor-signaal ALE.
- De geheugen adresbus is alleen op de basisprint (het zogenoemde "moederbord") aanwezig en deze vormt een

gemultiplexte versie van de systeem adresbus. Via deze bus worden achtereenvolgens de kolom- en rij-adressen van de geheugen-IC's naar het geheugen overgebracht.

- De X adresbus is via buffers gescheiden van de systeem adresbus en wordt gebruikt voor het adresseren van alle I/O-eenheden van het systeem.
  Ook de adressen van het ROM-BIOS (het meest fundamentele besturingssysteem), worden via deze bus verzorgd.

De databus

Ook de databus is gesplitst in vier delen.

- De lokale databus omvat de 16 datasignalen van de processor. Deze is bij de PC 8 bit breed en bij de XT 16 bit.
- De systeem databus wordt via latches op de lokale databus aangesloten.
  Ook deze is bij de PC 8 bit en bij de XT 16 bit breed.
- De geheugen databus is alleen op de hoofdkaart aanwezig en vormt de verbinding tussen het geheugen en de systeem databus.
- De X adresbus wordt via buffers met de systeem databus verbonden en wordt gebruikt voor het aanspreken van alle I/O-eenheden van het systeem.

De interrupt controller

Een belangrijk onderdeel in iedere computer is de interrupt controller. Deze beheert externe aanvragen voor gebruik van de processor, zoals bijvoorbeeld het toetsenbord dat aandacht van de processor opeist.

In de PC/XT-architectuur werd gebruik gemaakt van de 8259A, een PIC of programmeerbare interrupt controller. Dit IC heeft acht ingangskanalen. Iedere ingang is verbonden met een bepaalde perifere schakeling, die op deze manier dus onafhankelijk van elkaar een interrupt (onderbreking) bij de processor kunnen aanvragen.

Iedere interrupt heeft een benaming en een vast functie:

- IRQ0:
  interruptbenadering: 08_H
  functie: systeemtimer 8253
- IRQ1:
  interruptbenadering: 09_H
  functie: toetsenbord
- IRQ2:
  interruptbenadering: $0A_H$
  functie: gereserveerd
- IRQ3:
  interruptbenadering: $0B_H$
  functie: seriële poort 1
- IRQ4:
  interruptbenadering: $0C_H$
  functie: seriële poort 2
- IRQ5:
  interruptbenadering: $0D_H$
  functie: harde schijf controller
- IRQ6:
  interruptbenadering: $0E_H$
  functie: diskette controller
- IRQ7:
  interruptbenadering: $0F_H$
  functie: parallelle poort

De systeemtimer

In iedere PC/XT is een speciaal timer-IC aanwezig, de 8253. Deze PIT, oftewel programmeerbare interval timer, heeft drie onafhankelijk van elkaar programmeerbare tellers. Deze tellers worden als volgt gebruikt:

- eerste teller:
  systeemklok
- tweede teller:
  refresh van het geheugen
- derde teller:
  frequentie-besturing luidspreker

19.1 Architectuur van de computer hardware en de bus

Poortadres	Gebruik
000h-00fh	DMA-chip 8237A
020h-021h	PIC 8259A
040h-043h	PIT 8253
060h-063h	PPI 8255
080h-083h	DMA-paginaregister
0a0h-0afh	NMI-maskeringsregister
0c0h-0cfh	Gereserveerd
0e0h-0efh	Gereserveerd
100h-1ffh	Niet in gebruik
200h-20fh	Spelkaart
210h-217h	Uitbreidingseenheid
220h-24fh	Gereserveerd
278h-27fh	Parallelle printer
2f0h-2f7h	Gereserveerd
2f8h-2ffh	COM2
300h-31fh	Prototypekaart
320h-32fh	Schijfcontroller
378h-37fh	Parallelle interface
380h-38fh	SDLC-kaart
3a0h-3afh	Gereserveerd
3b0h-3bfh	Monochrome kaart/parallelle interface
3c0h-3cfh	EGA
3d0h-3dfh	Kleur-/grafische kaart
3e0h-3e7h	Gereserveerd
3f0h-3f7h	Diskettecontroller
3f8h-3ffh	COM1

Figuur 3/19.1-5: De gereserveerde adressen (poorten) voor bepaalde perifere schakelingen.

Het toetsenbord

Het toetsenbord wordt via de PPI 8255 met de systeembus verbonden. PPI staat voor programmeerbare periferie interface. Dit IC speelt ook bij het opstarten van het systeem een belangrijke rol. Via dit IC kan het BIOS bij het opstarten van het systeem alle systeemcomponenten testen. De 8255 bestuurt bovendien de luidspreker van het systeem AAN of UIT.

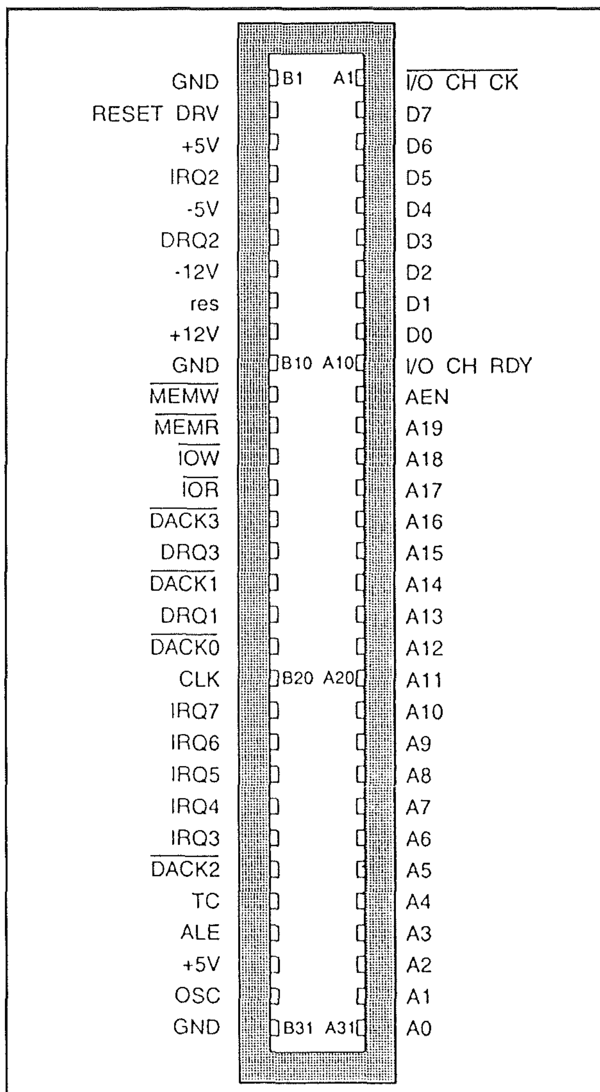
Gereserveerde poortadressen

Zoals reeds geschreven, worden perifere schakelingen via de adresbus geadresseerd. Dat wil zeggen dat deze schakelingen alleen met de bus worden verbonden (via tri-state transceivers) als hun adres op de adresbus staat. Deze adressen noemt men vandaar "poort-adressen", omdat zij als het ware poorten open zetten en sluiten tussen het systeem en de perifere schakelingen. Op deze manier wordt verhinderd dat twee perifere schakelingen op hetzelfde moment toegang tot de bus krijgen. Er zijn 26 adresbereiken gereserveerd voor intern gebruik. Fabrikanten van uitbreidingskaarten moeten er dus voor zorgen dat deze adresgebieden niet door hun kaart gebruikt worden. De gereserveerde adresgebieden en hun functie worden samengevat in de tabel van figuur 3/19.1-5.

De DMA-controller

Een belangrijke schakel in de architectuur van iedere PC is de DMA-controller. DMA staat voor "Direct Memory Access", directe toegang tot het geheugen. Via de DMA-controller is het mogelijk razend snel gegevens naar het geheugen te sturen, zonder dat de processor daar iets mee te maken heeft. De DMA-controller kan dus volledig zelfstandig het geheugen en de periferie benaderen.
In vaktermen zegt men dat de DMA-controller een ***"bus-master"*** is.
Hij kan immers, volledig zelfstandig, busbesturingen uitvoeren. In de PC/XT-architectuur zijn dus twee bus-masters aanwezig, de microprocessor zélf en de DMA-controller.
Beide bevinden zich op het moederbord, omdat de PC/XT-architectuur geen externe bus-masters tolereert. Het zou overigens duren tot bij de invoering van de

EISA- en MCA-architectuur alvorens PC's in staat waren externe bus-masters toe te laten in te dringen in de werking van het systeem. In de PC/XT-architectuur wordt als DMA-controller een 8237A toegepast.



Figuur 3/19.1-6: De aansluitgegevens van het slot van de PC/XT-architectuur.

De slot's van de PC/XT-architectuur

De systeembus staat ter beschikking van de uitbreidingskaarten via connectoren, die "bus-slot's" worden genoemd. Op de meeste moederborden uit de PC/XT-klasse heeft men zes slot's ingebouwd, waarvan de contacten parallel zijn geschakeld op de systeembus. Het maakt dus niet uit in welk slot men een uitbreidingskaart monteert. De aansluitgegevens van de PC/XT-slot's zijn getekend in figuur 3/19.1-6.

Pariteitslogica

Inleiding

In het blokschema van figuur 3/19.1-4 staat, boven het RAM-blokje, een klein blokje met daarin ingevuld het woord "pariteitslogica". De pariteitslogica maakt een belangrijk onderdeel uit van de architectuur van iedere PC en het is dus van belang er even bij stil te staan.

Iedere PC-gebruiker of -gebruikster zal wel eens ooit meegemaakt hebben dat het systeem vastloopt met een foutmelding:

```
Parity error
System halted
```

Men moet dan opnieuw opstarten en in de meeste gevallen is er verder niets aan de hand.

Die "Parity error" is een van de minst bekende en meest onbegrepen foutmeldingen die het systeem kan geven. En dat komt omdat men er, als alles goed is, maar zelden mee te maken krijgt. Toch is die foutmelding het resultaat van een PC-techniek, die zeer belangrijk is, maar voor de gemiddelde gebruiker totaal onbekend is: de "Parity Error Checking".

Controle op data-integriteit noodzakelijk

De man of vrouw voor het beeldscherm heeft, onbewust, een in feite ongelooflijk groot vertrouwen in de verzameling elektronische onderdelen waaruit een PC is samengesteld. Want die elektronische on-

derdelen moeten er voor zorgen dat gegevens, die via het toetsenbord worden ingevoerd, ook als dusdanig in het geheugen van het systeem terecht komen en vanuit dit geheugen worden overgebracht naar de harde schijf of naar een floppy. Nu is dát nog niet zo'n probleem. Maar men moet er ook zeker van zijn dat die eenmalig in het systeemgeheugen ingevoerde gegevens in de loop der uren niet veranderen. Met andere woorden: als men om negen uur 's ochtends een spreadsheet vol getallen in het geheugen intikt, dan moet men er zeker van zijn dat die gegevens om 17h00 nog steeds ongewijzigd in dat systeemgeheugen staan en bij het saven van het spreadsheet foutloos op de harde schijf terecht komen. En die eis is voor de elektronica heel wat moeilijker te realiseren!

Dynamische RAM's

Alle PC-geheugens zijn namelijk uitgevoerd onder de vorm van dynamische RAM's. De gegevens worden in deze IC's opgeslagen in een minuscuul condensatortje van minder dan 1 pF. De spanning over zo'n kleine condensator zal echter snel weglekken met als gevolg dat alle gegevens verloren zouden gaan als daar niet "iets" op gevonden was. Dat "iets" heet "refresh". De inhoud van het gehele systeemgeheugen wordt vele malen per seconde volledig ververst. Een ingewikkeld elektronisch proces, dat volledig bestuurd wordt door de processor. En het is voldoende dat er bij deze elektronische bewerkingen iets mis gaat met de voedingsspanning van het systeem, om gegevens in het geheugen te verminken. Het is bijvoorbeeld voldoende dat op een ongelukkig moment ergens in huis een koelkast inschakelt. Bij het inschakelen van een motor ontstaan stoorspanningen op het 220 V wisselspanningsnet. Deze smalle stoorpieken dringen door tot de systeemvoeding van de PC. Hoewel alle PC-voedingen voorzien zijn van uitgebreide onstoringsnetwerken kan het toch gebeuren dat zo'n smalle stoorpuls gedeeltelijk doordringt tot de +5 V voedingsspanning van de dynamische RAM's. En dan kan deze stoorpiek tot gevolg hebben dat de inhoud van één of meerdere geheugencellen verloren gaat.

Een IBM-idee

De aartsvader van de PC, IBM, heeft dit probleem onderkend en maatregelen in de PC-architectuur ingebouwd die er voor moeten zorgen dat de gegevensintegriteit van de in het systeemgeheugen opgeborgen gegevens gecontroleerd wordt. De techniek die daarvoor gebruikt wordt, wordt "Parity Error Checking" genoemd. Deze techniek is nadien (uiteraard) door alle kloon-fabrikanten overgenomen en is nu standaard in iedere architectuur aanwezig.
Dat geldt overigens niet voor alle computersystemen. De Macintosh-serie van Apple kent bijvoorbeeld geen controle op de data-integriteit van het geheugen.

De techniek van Parity Error Checking

Zoals algemeen bekend worden gegevens in de PC opgeslagen onder de vorm van bytes. Eén byte bestaat uit acht bits. Ieder bit kent slechts twee toestanden, namelijk "L" of "H".
Met deze 8 bits kan men in totaal 256 combinaties van "L"-en en "H"-en maken. Dat zijn de 256 verschillende gegevens die men in één byte kan opslaan.
In tegenstelling tot wat algemeen wordt aangenomen, worden deze 8 bit brede bytes achter niet in 8, maar in 9 geheugencellen opgenomen.

19.1 Architectuur van de computer hardware en de bus

DATA-BITS 0	1	2	3	4	5	6	7	AANTAL HOOG	PARITY-BIT
L	L	H	H	L	H	L	H	VIER	L
H	H	L	H	L	L	H	H	VIJF	H

Figuur 3/19.1-7: Het principe van pariteitscontrole.

De negende geheugencel wordt gebruikt als zogenoemd "parity-bit" en de inhoud van dit bit wordt gebruikt om de data-integriteit van het byte te controleren. IBM heeft een zeer eenvoudig systeem bedacht voor het controleren van de integriteit van een byte. Het systeem telt het aantal hoge bits in het byte. Is dit aantal even, dan wordt het parity-bit "L" gemaakt. Zijn er een oneven aantal "H"-bits in het byte, dan wordt het negende pariteits-bit "H" gemaakt. Het byte wordt nu, samen met het pariteits-bit in het systeemgeheugen opgeslagen. Deze techniek wordt toegelicht in de tabel van figuur 3/19.1-7. In het bovenste voorbeeld bestaat het byte uit vier hoge bits en wordt het parity-bit "L". In het onderste voorbeeld bestaat het byte uit vijf hoge bits, met als gevolg dat het parity-bit "H" wordt. Door de gekozen codering van het parity-bit zal een byte + parity-bit steeds een even aantal hoge bits bevatten! Bij het uit het systeemgeheugen halen van gegevens kan dit feit op een tamelijk eenvoudige manier elektronisch gecontroleerd worden. Leest het systeem een byte plus parity-bit uit het systeemgeheugen en stelt de elektronica vast dat er in die 9 bits een oneven aantal "H"-en zit, dan wordt dit doorgegeven aan het systeem en weet dit dat er iets fout is gegaan met de in het geheugen opgeslagen byte. Het zal duidelijk zijn dat het gekozen systeem niet alle mogelijke fouten kan opsporen. Zouden er in een byte twee bits van waarde veranderen, dan zal de sommatie van het aantal hoge bits toch kloppen en kan het systeem de fout niet ontdekken. Maar in ieder geval zorgt deze primitieve vorm van Parity Error Checking er toch voor dat de meeste geheugenfouten worden gedetecteerd.

Ingewikkelder systemen

Bij de moderne PC-architecturen is het systeem van pariteitscontrole behouden gebleven. Maar omdat men dan werkt met 16 en zelfs 32 bit brede bussen kan de in figuur 3/19.1-7 geschetste eenvoudige vorm niet gehandhaafd blijven. Dus heeft men ingewikkelder systemen ontwikkeld, die niet alleen gebruikt kunnen worden bij 8 bit brede databussen, maar ook bij 16 en 32 bit systemen. Bovendien leveren deze systemen betere prestaties. Het omwisselen van de waarde van twee bits wordt bijvoorbeeld zonder uitzondering gedetecteerd. Een van deze betere parity checkers is door Intel ontwikkeld en wordt bijvoorbeeld toegepast in de 8206. Deze met alle processoren van de 80-familie compatibele "Error Detection and Correction Unit" kan, zoals de naam het reeds zegt, niet alleen integriteitsfouten van het geheugen opsporen, maar deze ook corrigeren.

De 8206 kan:
- alle enkele bit fouten opsporen en corrigeren;
- alle dubbele bit fouten opsporen en corrigeren;
- de meeste meer-dan-twee bit fouten opsporen, maar niet corrigeren omdat deze worden geïnterpreteerd als enkel of dubbel bit fouten.

De 8206 werkt volgens het zogenoemde "Hamming-algoritme", een ingewikkelde wiskundige foutdetectie en -correctie techniek, die wel als groot nadeel heeft dat er veel meer dan één parity-bit noodzakelijk zijn. Om een indruk te geven is in de tabel van figuur 3/19.1-8 het aantal parity-bits gegeven dat de Hamming-methode nodig heeft voor databussen met verschillende breedte.

DATA-BUS BREEDTE	AANTAL HAMMING-BITS
8	5
16	6
24	6
32	7
40	7
48	8

Figuur 3/19.1-8: Het voor de Hamming-methode van pariteitscontrole noodzakelijk aantal bits bij verschillende data-breedtes.

De verwerking van de parity-check
Op de hoofdkaart van het systeem is een speciaal IC aanwezig, dat samen met de RAM-controller zorgt voor het coderen en nadien weer decoderen van het parity-bit. Als een parity-fout wordt vastgesteld vraagt dit IC een zogenoemde "Non Maskable Interrupt" (NMI) aan bij de processor. De interne structuur van de processor zorgt ervoor dat het systeem onmiddellijk stopt met alle werkzaamheden en met absolute prioriteit een bepaalde routine uit het BIOS gaat uitvoeren. Deze routine zet de foutmelding op het scherm en zet de processor nadien in een oneindig te doorlopen lus. Het gevolg is dat het systeem "hangt" en er niets anders op zit dan te herstarten.

De ISA- of AT-architectuur

Inleiding
Het tot op heden nog steeds meest gebruikte bus-systeem is de ISA-bus (Industry Standard Architecture) die door IBM met de AT (80286 processor) voor het eerst werd geïntroduceerd en derhalve ook als AT-bus bekend staat. Hierbij is sprake van een 16 bit brede databus. Er kunnen echter ook nog uitbreidingskaarten met een 8 bit brede bus worden gebruikt. Het ontwerp van de ISA-architectuur staat uiteraard niet op zich zelf. Deze architectuur is een verdere ontwikkeling van de allereerste PC/XT-architectuur. Vandaar dat het noodzakelijk is de ISA-architectuur in historisch perspectief te plaatsen.

Compatibiliteit
De techniek die bij het ontwerp van de architectuur van een PC is gebruikt, bepaalt in eerste instantie de geavanceerdheid en prestatiekenmerken van de PC. Bij het introduceren van de geavanceerde

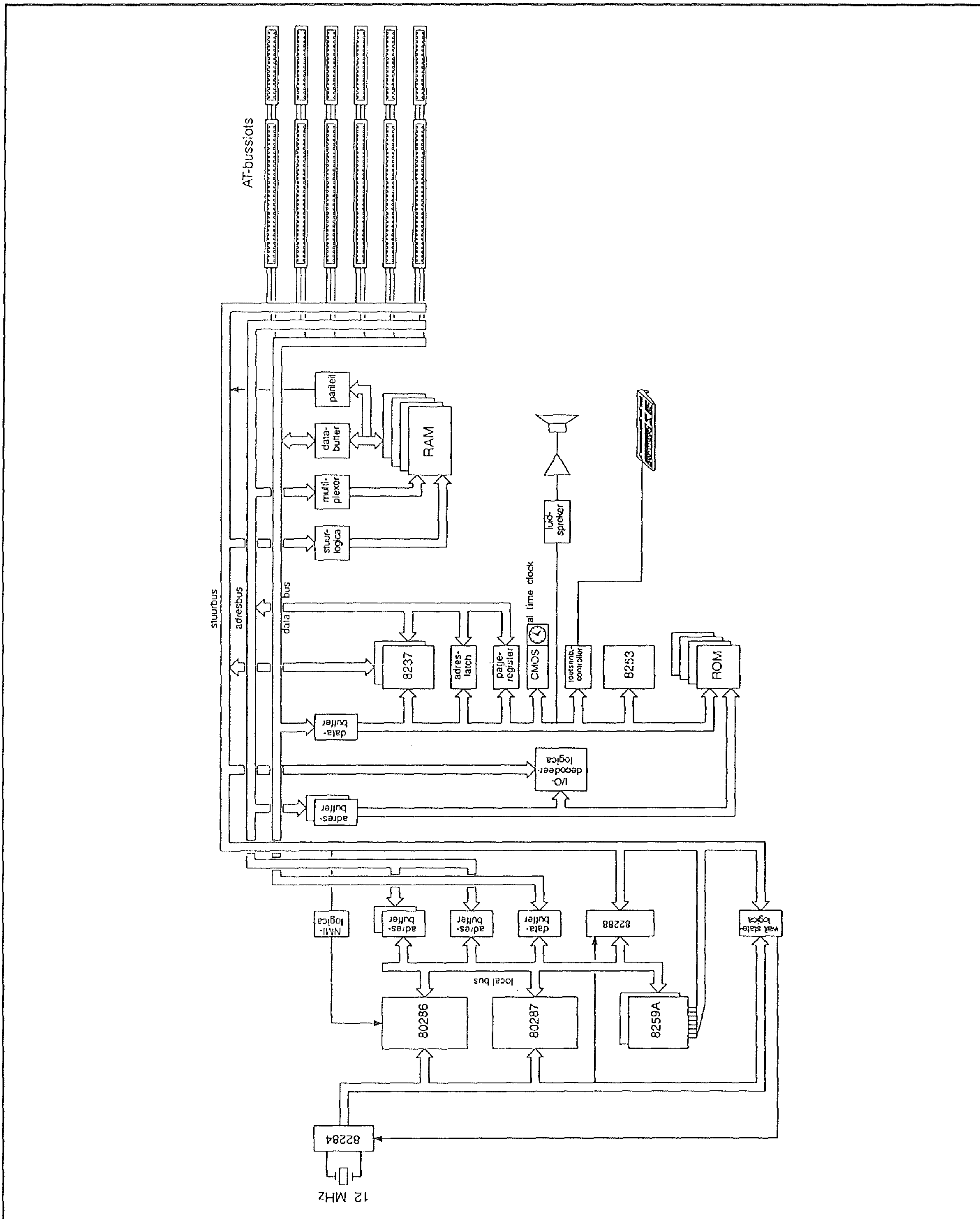
processor 80268 had men de mogelijkheid een zeer geavanceerde architectuur te ontwerpen. Maar zowel MicroSoft, IBM als Intel, de drie grote promotors van de PC, streefden er steeds naar compatibiliteit te garanderen. Bezitters van een PC/XT, uitgerust met een aantal uitbreidingskaarten, zouden deze kaarten zonder enig probleem ook in hun nieuw AT-systeem moeten kunnen gebruiken. Vandaar dat bij de invoering van de 80286 AT alleen werd gestreefd naar een uitbreiding van de oorspronkelijke PCXT-architectuur, zodat de flexibiliteit volledig bewaard bleef. De nieuwe 16 bit brede AT-architectuur bracht talrijke verbeteringen en een verhoging van de klokfrequentie, maar bleef neerwaarts compatibel met de oorspronkelijke PC/XT-architectuur, dat wil zeggen er was ook nog steeds plaats voor 8 bit brede uitbreidingskaarten.
Deze eigenschappen zorgden ervoor dat de AT-bus al gauw tot de industriële standaard werd verheven: de ISA-standaard. Deze standaard is tot op heden bepalend voor de compatibiliteit van de meeste PC-hardware.

Blokschema van een AT-systeem
Het algemeen blokschema van een computer die werkt volgens de AT- of ISA-architectuur is getekend in figuur 3/19.1-9. Hierbij moet wél opgemerkt worden dat dit blokschema de situatie van een allereerste generatie AT schetst. Nadien werden door diverse IC-fabrikanten zeer gecompliceerde chip's ontwikkeld, die de functies van een groot aantal van de in de figuur afzonderlijk weergegeven blokken in zicht verenigden. Maar voor het onderzoeken van de AT-technologie is de fundamentele ontrafeling van figuur 3/19.1-9 natuurlijk aanbevolen.

CMOS-geheugen
Een van de fundamentele nieuwigheden is de aanwezigheid van een klein CMOS-geheugen dat door een accu steeds op spanning wordt gehouden. In dit geheugen worden de fundamentele systeemparameters opgeslagen, zoals het type van harde schijf, de hoeveelheid geheugen, etc. Een tweede functie van dit altijd onder spanning staande deel van de architectuur is het bijhouden van de tijd en de datum. Hiervoor werd in eerste instantie een speciaal IC gebruikt, maar in latere versies werd geheugen- en kalender-elektronica in één chip geïntegreerd.
Dank zij dit geheugen hoeft men bij een AT geen schakelaartjes op het moederbord om te zetten om het systeem te configureren.
Dat kan softwarematig, via enige schermen waarin men bijvoorbeeld het soort harde schijf kan invullen. Bij het verlaten van dit zogenoemd SETUP-programma worden de ingevulde gegevens in het CMOS-geheugen opgeborgen.

De adresbus
Bij de AT-architectuur (en overigens bij alle opvolgers tot en met EISA) kan men vijf verschillende adresbussen onderscheiden.
- De lokale adresbus omvat de 24 adressignalen van de processor. Deze bus wordt door middel van latches van de rest van het systeem afgescheiden.
- De systeem adresbus is de belangrijkste bus van de architectuur. Zij bestaat uit de gelatchte lijnen A0 tot en met A19 van de lokale adresbus. Het vergrendelingssignaal is weer ALE.
- De geheugen adresbus is alleen op de moederkaart van het systeem aanwezig en vormt een gemultiplexte versie van de systeem adresbus.

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-9: De fundamentele architectuur van een AT-computer.

19.1 Architectuur van de computer hardware en de bus

Met de multiplexer worden in twee stappen de rij- en kolomadressen van de geheugenlocaties doorgegeven.
- De X adresbus is via buffers van de systeem adresbus gescheiden en verzorgt de toegang tot de I/O-unit's van het systeem. Dat zijn bijvoorbeeld de timer en een eventuele op de moederkaart aanwezige diskette-controller.
- Tot slot de L adresbus, die bestaat uit de adresbits A17 tot en met A23. Deze wordt uitgevoerd naar de slot's onder de vorm van de signalen LA17 tot en met LA23.

De databus
Deze bestaat, net zoals bij de PC/XT-architectuur, uit vier delen:
- de lokale bus;
- de systeembus;
- de geheugenbus;
- de X bus.

De functie van deze vier bussen is hetzelfde als bij de reeds besproken PC/XT-architectuur.

De hulpschakelingen
Ook de AT-architectuur bevat een aantal IC's die dienen voor het aan het werk zetten van het systeem. Zoals reeds geschreven, treft men deze chip's alleen bij de allereerste versies van het AT-systeem aan. Als snel werd de functionaliteit van deze chip's verenigd in maar een paar LSI-IC's.
Een overzicht van deze schakelingen:
- microprocessor (CPU), 80286;
- kloksignaalgever, 8284A;
- bus-controller, 82288;
- 2 x DMA-controller, 8287;
- 2 x interrupt-controller, 8259A;
- timer bouwsteen, 8254-2;
- real time clock, MC 146818;
- toetsenbord processor, 8042.

Verdere belangrijke componenten van een ISA-moederbord zijn een insteekvoet voor een wiskundige coprocessor en natuurlijk de insteekplaatsen voor uitbreidingskaarten (de zogenaamde slot's).

Twee interrupt-controllers
Zoals blijkt is een belangrijke vernieuwing ten opzichte van de PC/XT-architectuur de aanwezigheid van twee interrupt-controllers. Die worden respectievelijk master-PIC en slave-PIC genoemd. De beide schakelingen zijn in cascade geschakeld, met als gevolg dat er nu 16 interrupt-kanalen beschikbaar zijn.
Een overzicht van de toewijzing van deze interrupt's:
- IRQ0:
  interruptbenadering: 08_H
  functie: systeemtimer 8253
- IRQ1:
  interruptbenadering: 09_H
  functie: toetsenbord
- IRQ2:
  interruptbenadering: $0A_H$
  functie: slave-PIC
- IRQ3:
  interruptbenadering: $0B_H$
  functie: seriële poort 2
- IRQ4:
  interruptbenadering: $0C_H$
  functie: seriële poort 1
- IRQ5:
  interruptbenadering: $0D_H$
  functie: tweede parallelle poort
- IRQ6:
  interruptbenadering: $0E_H$
  functie: diskette controller
- IRQ7:
  interruptbenadering: $0F_H$
  functie: eerste parallelle poort
- IRQ8
  interruptbenadering: $0F_H$
  functie: real time klok

19.1 Architectuur van de computer hardware en de bus

Poortadres	Gebruik
000h-00fh	1e DMA-chip 8237A
020h-021h	1e PIC 8259A
040h-043h	PIT 8253
060h-063h	toetsenbordcontroller 8042
070h-071h	real-time klok
080h-083h	DMA-paginaregister
0a0h-0afh	2e PIC 8259A
0c0h-0cfh	2e DMA-chip 8237A
010h-0efh	gereserveerd
0f0h-0ffh	gereserveerd voor coprocessor 80287
100h-1ffh	vrij beschikbaar
200h-20fh	spelpoort-adapter
210h-217h	gereserveerd
220h-26fh	vrij beschikbaar
278h-27fh	2e parallelle interface
2b0h-2dfh	EGA
2f8h-2ffh	COM2
300h-31fh	prototypekaart
320h-32fh	vrij beschikbaar
378h-37fh	1e parallelle interface
380h-38fh	SDLC-adapter
3a0h-3afh	gereserveerd
3b0h-3bfh	monochroom-adapter/parallelle interface
3c0h-3cfh	EGA
3d0h-3dfh	color/graphics-adapter
3e0h-3e7h	gereserveerd
3f0h-3f7h	diskette-controller
3f8h-3ffh	COM1

Figuur 3/19.1-10: De toewijzing van de poortadressen aan de perifere schakelingen.

- IRQ9
  interruptbenadering: $0F_H$
  functie: omleiding naar IRQ2
- IRQ10
  interruptbenadering: $0F_H$
  functie: gereserveerd
- IRQ11
  interruptbenadering: $0F_H$
  functie: gereserveerd
- IRQ12
  interruptbenadering: $0F_H$
  functie: gereserveerd
- IRQ13
  interruptbenadering: $0F_H$
  functie: co-processor 80287
- IRQ14
  interruptbenadering: $0F_H$
  functie: harde schijf controller
- IRQ15
  interruptbenadering: $0F_H$
  functie: gereserveerd

Poortadressen

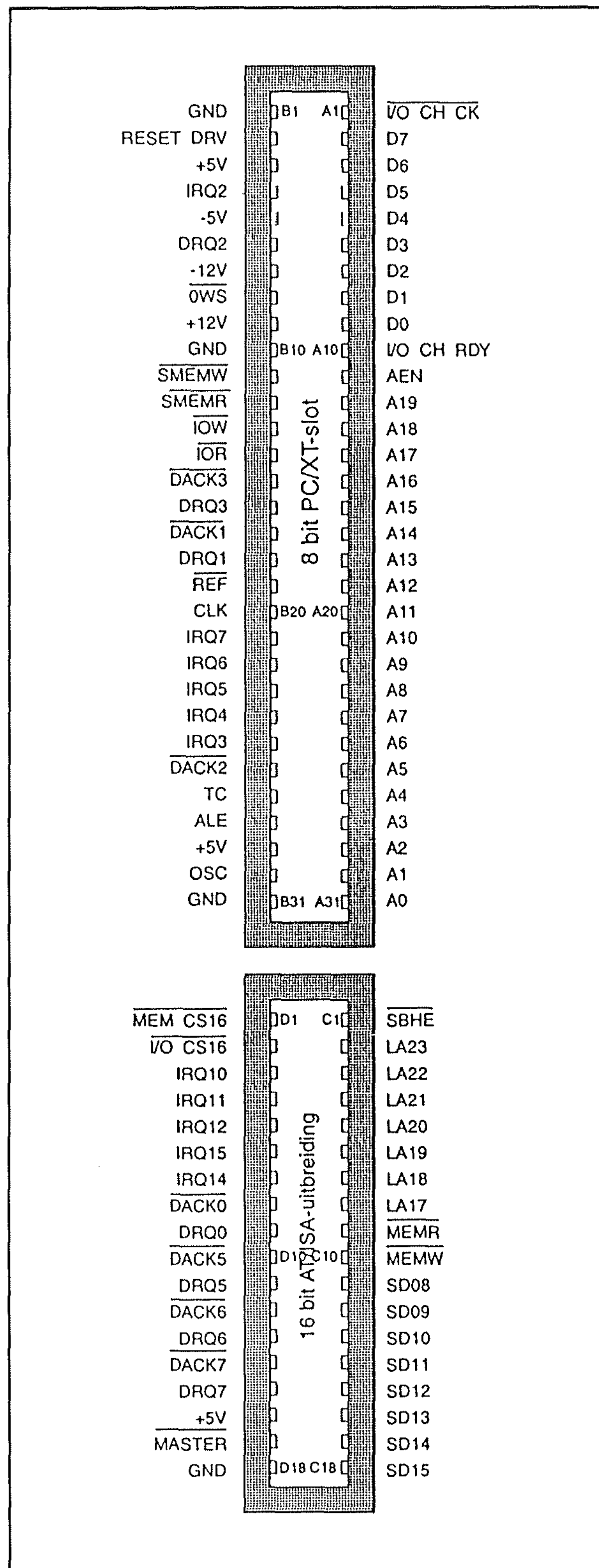
De AT-architectuur spreekt perifere apparatuur, net zoals de PC/XT, aan door middel van poortadressen. Omdat de AT-architectuur twee parallelle poorten ondersteunt en twee DMA's en PIC's aanwezig zijn, moest er wat gesleuteld worden aan de toewijzing van de beschikbare adressen.
In de tabel van figuur 3/19.1-10 wordt een volledig overzicht gegeven van de gestandaardiseerde toewijzing van alle poortadressen.

Het AT-slot

De insteekplaats (Engels: slot) van de acht bit brede PC-bus had 62 contacten. Via zo'n slot stonden op de PC-bus acht datalijnen, 20 adreslijnen en acht niet deelbare interruptlijnen ter beschikking. Een serieuze beperking was dat alle systeemcomponenten en uitbreidingskaarten deze acht interruptlijnen onder elkaar moesten verdelen zonder dat ze meervoudig konden worden gebruikt. De "zeggenschap" over de PC-bus uit 1981 berustte uitsluitend bij de centrale verwerkingseenheid (Engels: Central Processing Unit, CPU), met uitzondering van enkele situaties waarbij de DMA-controller (Direct Memory Access) het kortstondig voor het zeggen had.

(wordt vervolgd)

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-11: De aansluitgegevens van het AT-slot.

Met het beschikbaar komen van de 80286 microprocessor schiepen de ontwerpers bij IBM een nieuwe bus structuur. Daardoor werd een uitbreiding van de oorspronkelijke acht bit brede PC/XT-bus noodzakelijk. Het resultaat was dus de 16 bit brede "AT-bus" die als "ISA"-standaard de geschiedenis in zou gaan en de verdere ontwikkeling van de PC nadrukkelijk zou gaan beïnvloeden. Deze bus met de bijbehorende insteekplaatsen wordt tot op heden door alle fabrikanten van AT-kaarten gebruikt. Om de nieuwe mogelijkheden ook voor alle uitbreidingskaarten te ontsluiten, moest de busaansluiting tegenover de oorspronkelijke PC-bus worden uitgebreid. Voor dit doel werd een busuitbreiding met 38 lijnen ontwikkeld waardoor alle extra signalen voor het 16 bit gebruik konden worden aangestuurd. Door deze vorm van uitbreiding kon de oorspronkelijke PC-bus onveranderd nog aanwezig blijven en bleef een volledig neerwaartse compatibiliteit met 8 bit systemen gewaarborgd.
Zoals blijkt uit figuur 3/19.1-11 werden de nieuwe signalen via een extra connector aan de uitbreidingskaarten aangeboden. Op deze manier konden de oude 8 bit brede kaarten zonder problemen gebruikt blijven worden en konden nieuw ontwikkelde 16 bit brede kaarten van de volledige functionaliteit van de AT-architectuur gebruik maken.

AT versus ISA

In de inleiding werd gesteld dat de benamingen AT-architectuur en ISA-architectuur identiek zijn. Dat is in feite niet exact! De naam AT is een typische IBM-benaming en was oorspronkelijk allen bedoeld voor intern IBM-gebruik. Maar IBM heeft de specificaties van de AT-bus nooit duidelijk op papier gezet.

Toen echter de technologie snel verder ontwikkelde en allerlei fabrikanten klonen van AT-systemen op de markt brachten met snellere klokken, ontstonden er compatibiliteitsproblemen.
Om deze problemen op te lossen werd, zeven jaar na de introductie van het eerste AT-systeem, een internationale vervanging voor de AT-standaard opgesteld. De bedoeling was dat alle fabrikanten die iets met computer-technologie te maken hadden, zich aan deze standaard zouden houden. Deze standaard werd ISA genoemd, afkorting van "Industrial Standard Architecture". Deze ISA-standaard is voor ongeveer 95 % volledig compatibel met de door IBM gebruikte AT-architectuur.

Beperkingen van de ISA-architectuur
Met de verdere ontwikkeling van uitbreidingskaarten, de gestegen eisen, hogere klokfrequenties en de invoering van de 32 bit processor 80386, kwamen echter de tekortkomingen en de beperkingen van de ISA-architectuur meer en meer boven water. Langs verscheidene wegen trachtte men in eerste instantie oplossingen te vinden voor deze onvolkomenheden. Ondanks alle goed bedoelde pogingen is het toch niet gelukt alle nadelen en beperkingen van de ISA-bus techniek te overkomen. Vooral de volgende beperkingen bleven hinderlijk bestaan:
- lage gegevensoverdracht snelheid van maximaal 5 MB/s;
- geen onderling deelbare interrupts;
- alleen de processor en de DMA-controller kunnen de bus "controleren";
- geen 32 bit data-bus (alleen 16 bit);
- alleen 24 adreslijnen (vandaar slechts maximaal 16 MB adresbereik);
- de DMA-controller kan alleen 64 kB blokken overdragen.

Dit overzicht maakt duidelijk dat een snelle 32 bit processor door het 16 bit corset van de ISA-architectuur behoorlijk wordt geremd.

Beroemde AT chip-set's
Terwijl bij de moederborden van het eerste uur alle functies door afzonderlijke bouwstenen werden uitgevoerd, werden vrij snel complete chip-set's aangeboden die de functies van een of meer bouwstenen in zich verenigen.
Wezenlijke voordelen hiervan zijn het gebruik van zogenaamde "Shadow-RAM" functies, de mogelijkheid om het intern geheugen op het moederbord uit te breiden, individuele klokinstellingen, het zogenaamde "Paging" (een speciale toegangsmethode tot het interne geheugen), lager stroomverbruik en de hogere prestaties van het moederbord bij kleinere afmetingen. Met het oog op de compatibiliteit moesten echter de functionele en operationele kenmerken van de oorspronkelijke bouwstenen behouden blijven. De eerste chipset's beperkten zich trouwens nog tot de 80286; ze boden nog geen mogelijkheden voor gedetailleerde instellingen via een uitgebreide setup procedure zoals die vandaag de dag standaard zijn.
Een van de eerste en ook de meest bekende chipset's in dit verband was de "NEAT" chipset (van de firma Chips & Technologies) die bij veel verschillende AT-typen werd ingezet. De NEAT chipset bestaat bij de 80286 en bij de 80386SX PC uit vier en bij de 80386 uit zeven chip's. In alle gevallen is als zogenaamde periferie controller de bouwsteen 82C206 aanwezig. De overige chip's zijn weliswaar verschillend en op de specifieke eisen van de microprocessor afgestemd, voor de gebruiker onderscheiden ze zich qua werking echter nauwe-

lijks. De NEAT en vergelijkbare chipset's zijn voornamelijk door de veelzijdige mogelijkheden voor het programmeren en instellen van interne registers bijzonder aantrekkelijk voor zowel PC-fabrikanten als (gevorderde) gebruikers. Voor de programmeur zijn er tal van mogelijkheden de registers van de chip's op eenvoudige wijze aan te spreken. Via de SETUP-procedure kan echter ook de gebruiker de eigenschappen optimaal afstemmen op zijn systeem.
Belangrijke instellingen die men als gebruiker via de SETUP kan wijzigen, betreffen de klokfrequentie van het systeem en van verscheidene onderdelen, zoals de microprocessor, de bus en de in- en uitvoerkanalen. Men kan daarbij kiezen uit vier verschillende modi. Andere belangrijke instellingen hebben te maken met de vaststelling van de geheugentoegang, bijvoorbeeld wat betreft de "Wait States" (nul of meerdere klokpulsen wachten met het ondernemen van een verdere actie) voor het lezen en schrijven in het geheugen. Daarnaast zijn er nog diverse cache-opties.
Geoptimaliseerde instellingen kunnen enerzijds de prestaties van een PC in hoge mate verbeteren, anderzijds kunnen verkeerde instellingen ook de werking of betrouwbaarheid van een PC -systeem nadelig beïnvloeden.

De Micro Channel Architecture

Inleiding

Met de invoering van de 32 bit microprocessoren werd in toenemende mate de noodzaak duidelijk een nieuw bussysteem en daarmee ook een nieuwe moederbord architectuur te ontwikkelen. Een 16 bit brede bus met tal van ongebruikte mogelijkheden en een beperkt adresbereik zoals bij de ISA-architectuur was eenvoudigweg niet meer te rijmen met de mogelijkheden van snelle 32 bit processoren.
Weer was het IBM die, ook wel uit concurrentieoverwegingen, in 1987 met een volstrekt nieuwe architectuur op de proppen kwam, de zogenoemde "Micro Channel Architecture", afgekort MCA. De MCA-architectuur werd in eerste instantie voor de 80286-processor met een 16 bit brede bus en in een uitgebreide vorm voor 80386- en 80486-processoren met een 32 bit bus gerealiseerd. Enige rol van betekenis heeft echter alleen de 32 bit variant van dit alternatief voor de ISA-architectuur gespeeld. Kenmerkend voor de MCA-architectuur zijn een verbeterde moederbord architectuur, aanzienlijke verhoging van de geheugencapaciteit en rekenprestaties alsmede een optimalisering van de systeemcomponenten.

VGA

Tegelijk met de invoering van het MCA-concept werd door IBM eveneens een nieuwe grafische standaard voor beelschermen in het leven geroepen, de (intussen al weer verbeterde) VGA-standaard (Video Graphics Array).

Geen nieuwe standaard

Hoewel de MCA-architectuur tegenover de oude ISA-standaard een keur van voordelen biedt, is deze architectuur altijd onderwerp van diskussie geweest. Het is dan ook geen internationaal geaccepteerde standaard geworden! Voornaamste redenen daarvoor zijn vermoedelijk de incompatibiliteit en de totale breuk met de ISA-architectuur, waardoor de talloze in ge-

bruik zijnde ISA uitbreidingskaarten voor een MCA-PC volledig onbruikbaar zijn. In tegenstelling tot andere bussystemen is de MCA-techniek bovendien gepatenteerd zodat ze niet nagebouwd kan worden zonder dat aan IBM licentierechten moeten worden afgedragen. Dit is onder meer ook een reden waarom PC-fabrikanten nooit hebben staan te popelen om MCA-systemen te produceren.

Voordeel

Een van de grootste voordelen van de nieuwe systeemarchitectuur voor 80386-processoren en hoger is ongetwijfeld het doorbreken van de 16 bit beperking. De 32 bit variant van MCA biedt zowel een 32 bit databus al een 32 bit adresbus. Resultaat: een aanzienlijk snellere gegevensoverdracht en een bijna onvoorstelbaar groot adresbereik van 4 GB (4.000 MB). Met enig kunst- en vliegwerk behaalde de oude ISA-bus een overdrachtsnelheid van ongeveer 5 MB/s, MCA bereikt gemakkelijk het dubbele (10 MB/s).

Belangrijkste noviteit: de Bus Master

De eigenlijke prestatieverbetering dankt MCA echter aan de invoering van de zogenoemde "Bus Master". Bij de ISA-bus is de CPU de alles overheersende "meester" in het systeem. Dientengevolge vindt busgebruik altijd via de centrale microprocessor plaats. Bij MCA is dat anders geregeld: door gebruik te maken van een "Bus Master Chip Set" kunnen meer microprocessoren of controllers tot het systeem worden toegelaten. Dankzij dit feit kunnen ook intelligente uitbreidingskaarten die een eigen microprocessor bezitten, in het systeem functioneren en over de volle breedte van de 32 bit MCA-bus beschikken. De microprocessor op zo'n uitbreidingskaart functioneert als "Slave" controller en is op deze manier in staat processen te realiseren die onafhankelijk van het overige PC-systeem aflopen. De bus master zorgt daarbij binnen het systeem voor de coördinatie van MCA-bus, van de extra microprocessor op de uitbreidingskaart (voor slave-functies) en van de eigenlijke centrale microprocessor van het systeem (voor de masterfuncties). In de bus master opzet werken de extra processoren dus concurrerend met de centrale processor op het moederbord en alle I/O-bouwstenen met een eigen microprocessor kunnen rechtstreeks gebruik maken van het gemeenschappelijke bus-systeem. Voor dit doel is onder meer ook de interrupt aansturing gewijzigd.

Cache-technologie

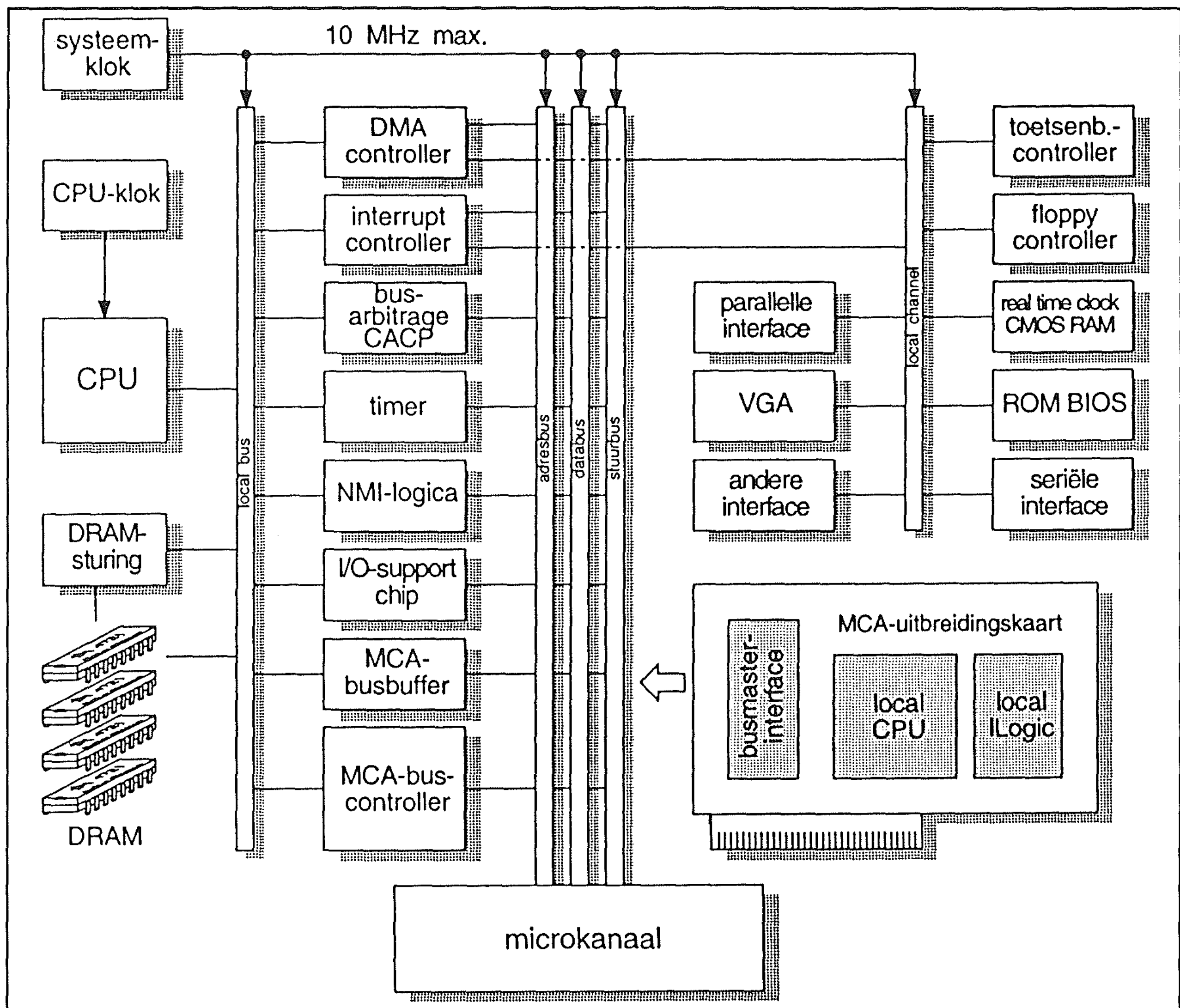
Een ander voordeel van de MCA-architectuur is de verbeterde toegang tot gegevens met behulp van de zogenaamde cache-techniek. Een cache-controller (82385) en cache-geheugen van minimaal 32 kB, uitgevoerd met bijzonder snelle geheugenchips, zorgen ervoor dat gegevens uit het geheugen vele malen sneller toegankelijk zijn.

Een van de beperkingen van de ISA-bus, namelijk het feit dat de DMA-controller alleen blokken van maximaal 64 kB kan doorgeven, behoort bij een MCA-systeem eveneens tot het verleden.

Chipset's voor MCA

Bij de MCA-technologie komen als chipset's op het moederbord alleen in hoge mate geïntegreerde VLSI-chip's (Very Large Scale Integration) in aanmerking. Omdat deze bouwstenen op de systeemeisen zijn afgestemd, bevatten ze reeds een aantal systeemfuncties die voordien met uitbreidingskaarten gerealiseerd moesten worden.

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-12: Het blokschema van de MCA-architectuur.

Zo zijn bijvoorbeeld de parallelle en seriële interface alsmede een VGA grafische beeldschermadapter reeds in het moederbord geïntegreerd. Daardoor zijn niet alleen minder uitbreidingskaarten nodig, maar neemt ook de betrouwbaarheid van het systeem toe.

Automatische configuratie

Een ander vooral voor de gebruiker aardige bijkomstigheid zijn de intelligente uitbreidingskaarten. Dankzij individuele identificaties hoeven deze niet meer uitdrukkelijk aan het systeem bekend te worden gemaakt, bijvoorbeeld door zogenaamde DIP-switches of via setup procedures. Het MCA-bus systeem herkent de uitbreidingen uit zichzelf en zorgt voor automatische configuratie. Een praktisch handvat vergemakkelijkt bovendien het plaatsen van de kaart in een vrij slot op het MCA moederbord.
Nadeel hiervan is echter dat alleen MCA-uitbreidingen kunnen worden ingebouwd die of van IBM zélf of van een licentienemer afkomstig zijn die daarvoor

aan IBM moet betalen. De afgelopen jaren hebben geleerd dat niet al te veel hardware producenten daar happig op zijn.

Blokschema van de MCA-architectuur
In figuur 3/19.1-12 is het blokschema van een typisch MCA-systeem getekend. Een belangrijk gegeven is dat er twee klokgeneratoren aanwezig zijn. Op de eerste plaats de afzonderlijke systeemklok, die alle onderdelen van de architectuur voorziet met een klok van 10 MHz. Op de tweede plaats de CPU-klok, die de processor stuurt met een aan de maximale werkfrequentie van de processor aangepaste klok. De lokale geheugenbus tussen de processor en het geheugen wordt ook door deze klok gestuurd. Op deze manier kan men de gegevensoverdracht tussen processor en geheugen aanpassen aan de eigen klokfrequentie van de processor.

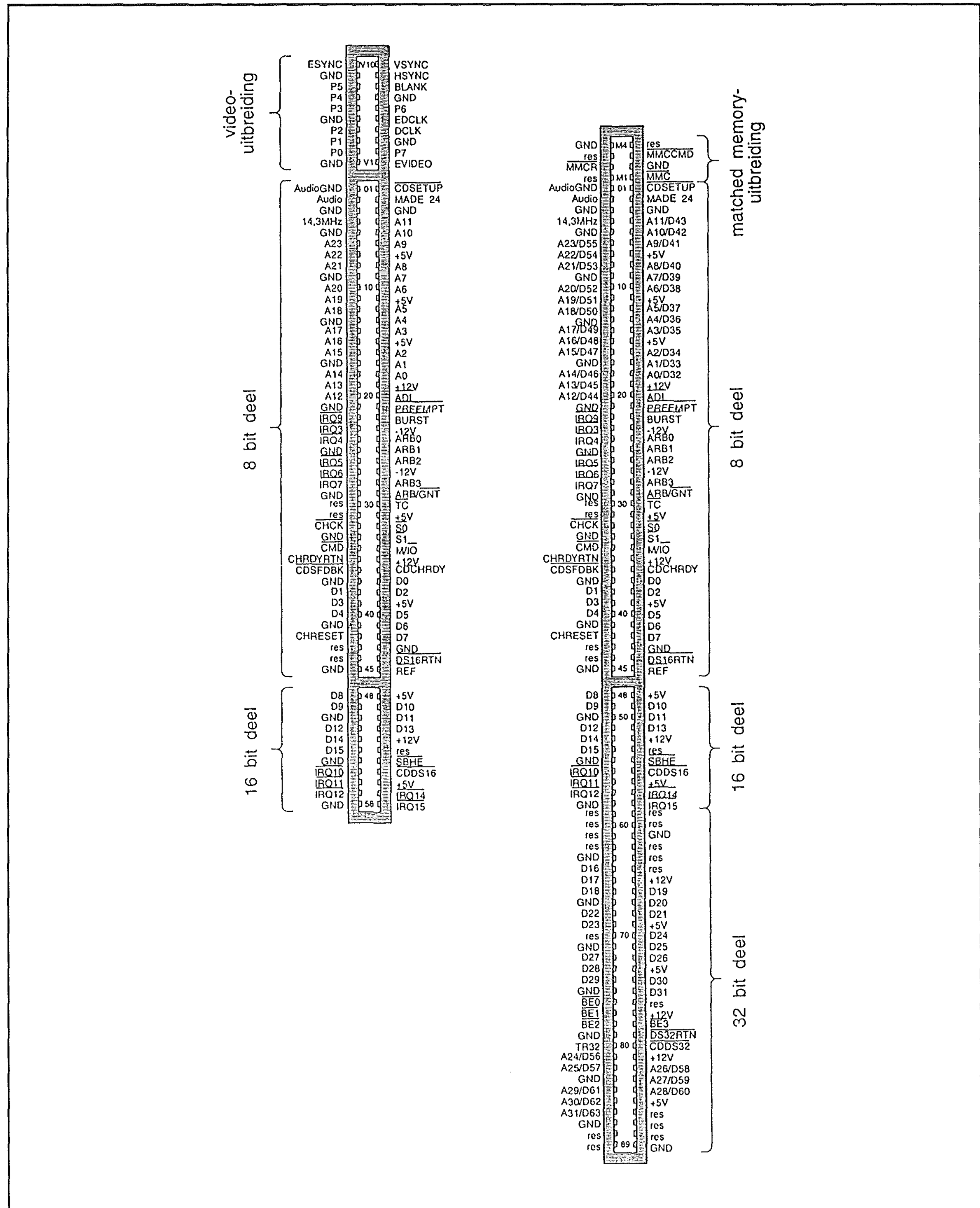
I/O-poorten
Ook bij de MCA-architectuur worden de perifere bouwstenen via poorten benaderd. In de tabel van figuur 3/19.1-13 wordt een overzicht gegeven van het I/O adresbereik van MCA.

Het MCA-slot
Omdat IBM iedere compatibiliteitsgedachte overboord had gezet, hadden de ontwerpers van de MCA-architectuur alle vrijheid om en volledig nieuw slot te ontwerpen. Wél moest men er rekening mee houden, dat er uitbreidingskaarten op de markt zouden komen met breedtes van 8, 16 en 32 bit. Het is natuurlijk onzin om een 32 bit kaart te ontwerpen voor het aansturen van een printer! De kaart zou alleen maar extreem duur worden, zonder dat er enige extra functionaliteit aanwezig zou zijn.

I/O-adres	Betekenis
000h-01fh	master-DMA
020h-021h	master-8259A
040h-043h	timer 1
044h-047h	failsafe-timer
060h-064h	toetsenbord/muis-controller
070h-071h	real-time clock/CMOS-RAM
074h-076h	uitgebreid CMOS-RAM
080h-08fh	DMA-paginaregister
090h	CACP
091h	feedback-register
092h	systeemstuurpoort A
096h	adapteractivering/setup-register
0a0h-0afh	slave-8259A
0c0h-0dfh	slave-DMA
0e0h-0efh	geheugenconfiguratieregister
0f0h-0ffh	coprocessor
100h-107h	POS-register 0-7
200h-20fh	spelpoort-adapter
210h-217h	gereserveerd
220h-26fh	vrij beschikbaar
278h-27fh	2e parallelle interface
2b0h-2dfh	EGA
2f8h-2ffh	COM2
300h-31fh	prototypekaart
320h-32fh	vrij beschikbaar
378h-37fh	1e parallelle interface
380h-38fh	SDLC-adapter
3a0h-3afh	gereserveerd
3b0h-3bfh	monochroom-adapter
3b4h-3bah	VGA
3bch-3beh	parallelle interface
3c0h-3dfh	EGA/VGA
3e0h-3e7h	gereserveerd
3f0h-3f7h	diskette-controller
3f8h-3ffh	COM1

Figuur 3/19.1-13: De I/O-poorten van de MCA-architectuur en hun toewijzing.

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-14: Een volledig overzicht van alle contacten van het MCA-slot.

Vandaar dat het hart van het MCA-slot wordt gevormd door een 8 bit sectie met 90 contacten. Deze contacten liggen in één vlak en staan heel dicht bij elkaar. Naast deze 8 bit sectie heeft men uitbreidingen tot 16 bit en 32 bit. In totaal heeft het gehele MCA-slot niet minder dan 202 contacten, maar dank zij de kleine afstand tussen de contacten is een MCA-slot toch heel compact.
Een volledig overzicht van alle contacten van een MCA-slot is getekend in figuur 3/19.1-14.

De EISA-architectuur

Inleiding
Hoewel de MCA-techniek in principe met alle beperkingen en nadelen van de ISA-standaard heeft afgerekend, is het nooit tot een echte doorbraak van deze IBM-vinding gekomen. Ze wordt bijna uitsluitend toegepast in IBM PC's van het type PS/2. Reden daarvoor zijn vooral het genoemde licentiebeleid van IBM en de volstrekte onverenigbaarheid met bestaande ISA-kaarten.
Het gevolg hiervan was dat enkele grote PC-fabrikanten (zonder IBM) de koppen bij elkaar staken om een alternatief voor het MCA-concept te ontwikkelen. Deelnemers waren grote namen op het PC-toneel als Compaq, Olivetti, Epson, NEC, HP, Tandy, Wyse en AST. Doel van de gezamenlijke inspanningen was de ontwikkeling van een nieuwe busarchitectuur die volledig rekening hield met de eisen van een 32 bit microprocessor, maar tegelijkertijd ook neerwaarts compatibel was met de oude ISA-architectuur. Het resultaat werd de zogenaamde "Extended Industry Standard Architecture", EISA gedoopt, een verdere ontwikkeling op de basis van de oude ISA-standaard. Het EISA-systeem accepteert zowel nieuwe EISA- als ook de voorheen gebruikelijke ISA-kaarten.

Open bus systeem
EISA is een open bus systeem met uitstekende eigenschappen dat zowel aan de eisen van de nieuwe microprocessoren voldoet met 32 datalijnen op de EISA-bus als geschikt is voor 16 of zelfs 8 bit busgebruik volgens ISA-normen. EISA is zodoende geen nieuw concept, maar een logische verbetering en uitbreiding van het oude ISA-concept met het oog op een 32 bit data bus.

EISA contra MCA
Het in zes of zeven lagen vervaardigde EISA moederbord beschikt ook over de meeste innovaties en voordelen van de eerder besproken MCA-techniek. Niet alleen de databus is 32 bit breed, maar ook de adresbus (adresbereik tot 4 GB). De interne doorvoersnelheid van gegevens is met 33 MB/s in de 32 bit modus vergelijkbaar met die van MCA. Voor het cachegeheugen en de cache-controller worden dezelfde (of vergelijkbare) bouwstenen gebruikt. Dat EISA ook verscheidene bus masters toelaat, zal dan niemand meer verbazen.
Uitbreidingskaarten worden automatisch geconfigureerd. Het systeem herkent bijvoorbeeld of het met een 32 bit EISA-kaart te maken heeft of "slechts" met een ISA-kaart. Hierdoor is gewaarborgd dat alle bestaande ISA-kaarten zonder haperen op een EISA moederbord blijven werken. De overstap op de duurdere EISA-kaarten kan zodoende stap voor stap gebeuren. Tot volledige ontplooiing komt zo'n EISA moederbord uiteraard alleen met echte

19.1 Architectuur van de computer hardware en de bus

32 bit uitbreidingskaarten en zonder te worden belemmerd door de aanwezigheid van een of meerdere "ouderwetse" 16 bit kaarten. Bijzondere voordelen doen zich voor bij het gebruik van op de EISA-technologie geënte disk-controllers, grafische kaarten en netwerkadapters.

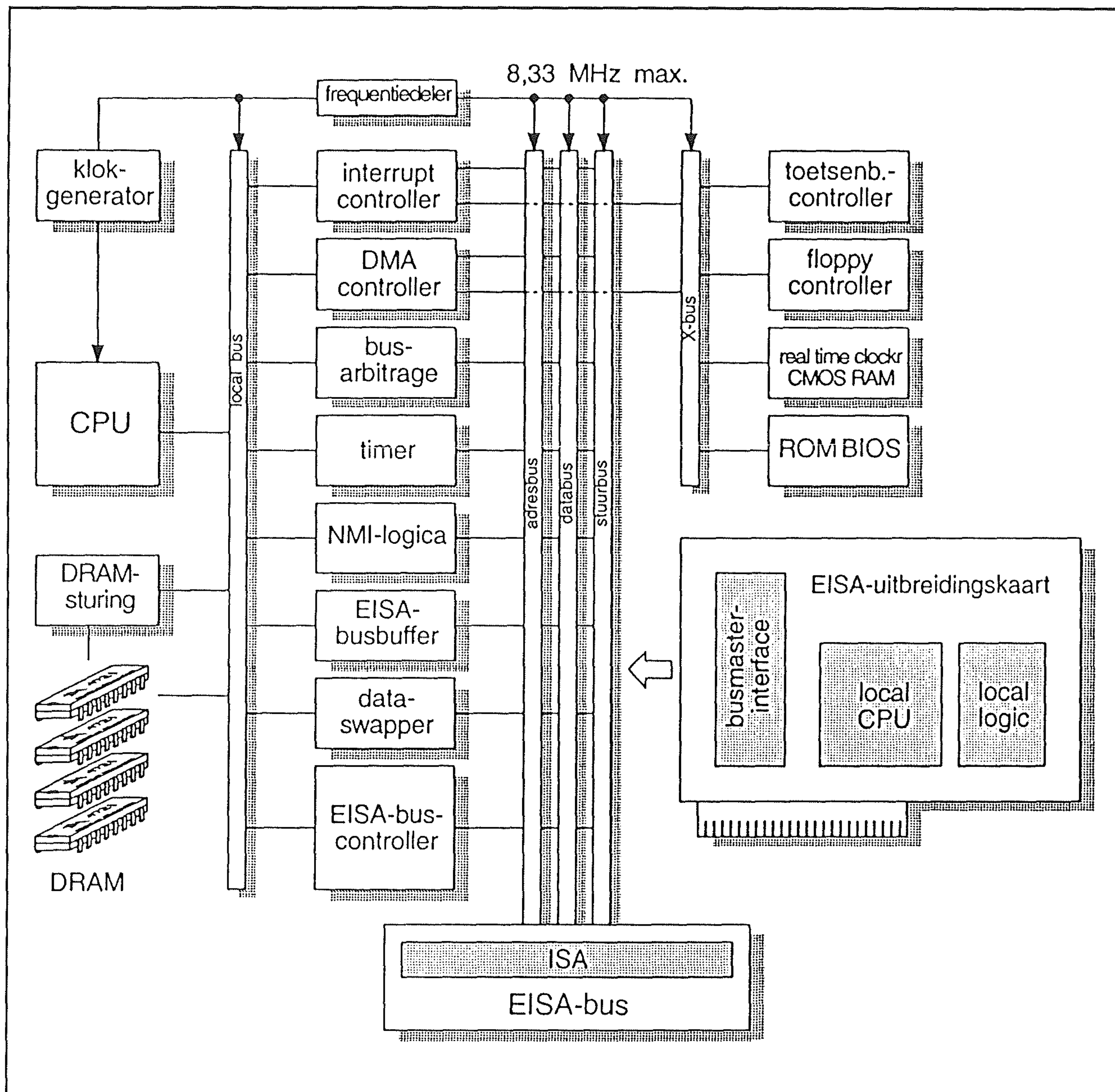
Compatibiliteit heeft ook nadelen
Neerwaartse compatibiliteit is mooi, maar heeft natuurlijk ook nadelen. Dat geldt nu voor Windows 95, dat vanwege die neerwaartse compatibiliteit toch nog volledig op MS-DOS leunt, dat gold dus ook toen men overstapte van ISA naar EISA.
Een compromis moest bijvoorbeeld worden bereikt ten aanzien van de interrupt-aansturing die op twee verschillende manieren is geregeld. 32 bit uitbreidingskaarten kunnen een van de weinige hardware-interrupts met elkaar delen, een ISA-kaart daarentegen eist zo'n interruptlijn alleen en uitsluitend voor zichzelf op. Daaruit volgt nog een ander nadeel als EISA- en ISA-kaarten in hetzelfde adresbereik liggen. Omdat het in dat geval niet mogelijk is twee "modi" gelijktijdig naast elkaar te bedrijven wordt ook de 32 bit EISA-kaart veroordeeld tot de langzame 16 bit modus.

Blokschema van de EISA-architectuur
In figuur 3/19.1-15 is het blokschema van een typisch EISA-systeem getekend. De klokgenerator bedient nu zowel de processor als de EISA-bus, maar tussen processoraansturing en bus is en frequentiedeler tussengeschakeld. De EISA-architectuur werkt dus volledig synchroon, omdat het gehele systeem uit één klokgenerator wordt gestuurd. De maximale frequentie van een EISA-bus is nog steeds 8,33 MHz. De processor kan echter via de locale bus het geheugen op een hogere snelheid adresseren. Het hart van een EISA-systeem is de EISA-controller. Deze onderscheidt EISA van ISA busbenaderingen. Hij verdeelt 32 bit transporten, indien noodzakelijk, in 8 of 16 bit stukken voor 8 of 16 bit randapparatuur.

Het interrupt-systeem
EISA kent ook 16 interrupt-niveaus, die door één interrupt-controller worden bestuurd. In tegenstelling tot bij de ISA-architectuur is een interrupt-kanaal niet meer exclusief voor één randapparaat gereserveerd.
Een overzicht van de toewijzing van de 16 interrupt-kanalen:

- IRQ0:
  prioriteit: 1
  interruptbenadering: 08_H
  functie: systeemtimer 8253
- IRQ1:
  prioriteit: 2
  interruptbenadering: 09_H
  functie: toetsenbord
- IRQ2:
  prioriteit: -
  interruptbenadering: $0A_H$
  functie: slave-PIC
- IRQ3:
  prioriteit: 11
  interruptbenadering: 73_H
  functie: seriële poort 2
- IRQ4:
  prioriteit: 12
  interruptbenadering: 74_H
  functie: seriële poort 1
- IRQ5:
  prioriteit: 13
  interruptbenadering: 75_H
  functie: tweede parallelle poort
- IRQ6:
  prioriteit: 14
  interruptbenadering: 76_H
  functie: diskette controller

19.1 Architectuur van de computer hardware en de bus



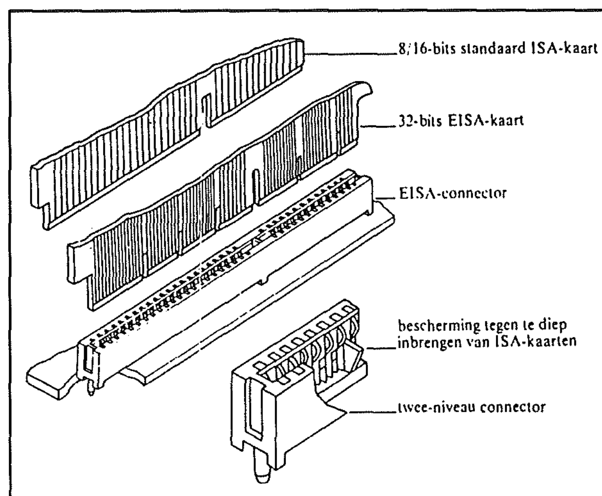
Figuur 3/19.1-15: Blokschema van de EISA-architectuur.

- IRQ7:
  prioriteit: 15
  interruptbenadering: 77_H
  functie: eerste parallelle poort
- IRQ8
  prioriteit: 3
  interruptbenadering: $0B_H$
  functie: real time klok
- IRQ9
  prioriteit: 4
  interruptbenadering: $0C_H$
  functie: vrij beschikbaar
- IRQ10
  prioriteit: 5
  interruptbenadering: $0D_H$
  functie: vrij beschikbaar

19.1 Architectuur van de computer hardware en de bus

I/O-adres	Betekenis
0000h...00ffh	ISA-moederbord
0100h...3fffh	ISA-uitbreidingskaarten
0400h...04ffh	gereserveerd voor EISA-moederbord-controller
0800h...08ffh	gereserveerd voor EISA-moederbord
0c00h...0cffh	gereserveerd voor EISA-moederbord
1000h...1fffh	uitbreidingsslot 1
2000h...2fffh	uitbreidingsslot 2
3000h...3fffh	uitbreidingsslot 3
4000h...4fffh	uitbreidingsslot 4
5000h...5fffh	uitbreidingsslot 5
6000h...6fffh	uitbreidingsslot 6
7000h...7fffh	uitbreidingsslot 7
8000h...8fffh	uitbreidingsslot 8
9000h...9fffh	gereserveerd voor extra uitbreidingsslots

Figuur 3/19.1-16: De toekenning van de I/O-poortadressen bij de EISA-architectuur.



Figuur 3/19.1-17: Bij de EISA-connector werden contacten op twee niveaus aangebracht, waardoor zowel oude ISA-kaarten als moderne EISA-kaarten zonder problemen passen.

- IRQ11
  prioriteit: 6
  interruptbenadering: $0E_H$
  functie: vrij beschikbaar
- IRQ12
  prioriteit: 7
  interruptbenadering: $0F_H$
  functie: vrij beschikbaar
- IRQ13
  prioriteit: 8
  interruptbenadering: 70_H
  functie: co-processor 80387
- IRQ14
  prioriteit: 9
  interruptbenadering: 71_H
  functie: harde schijf controller
- IRQ15
  prioriteit: 10
  interruptbenadering: 72_H
  functie: vrij beschikbaar

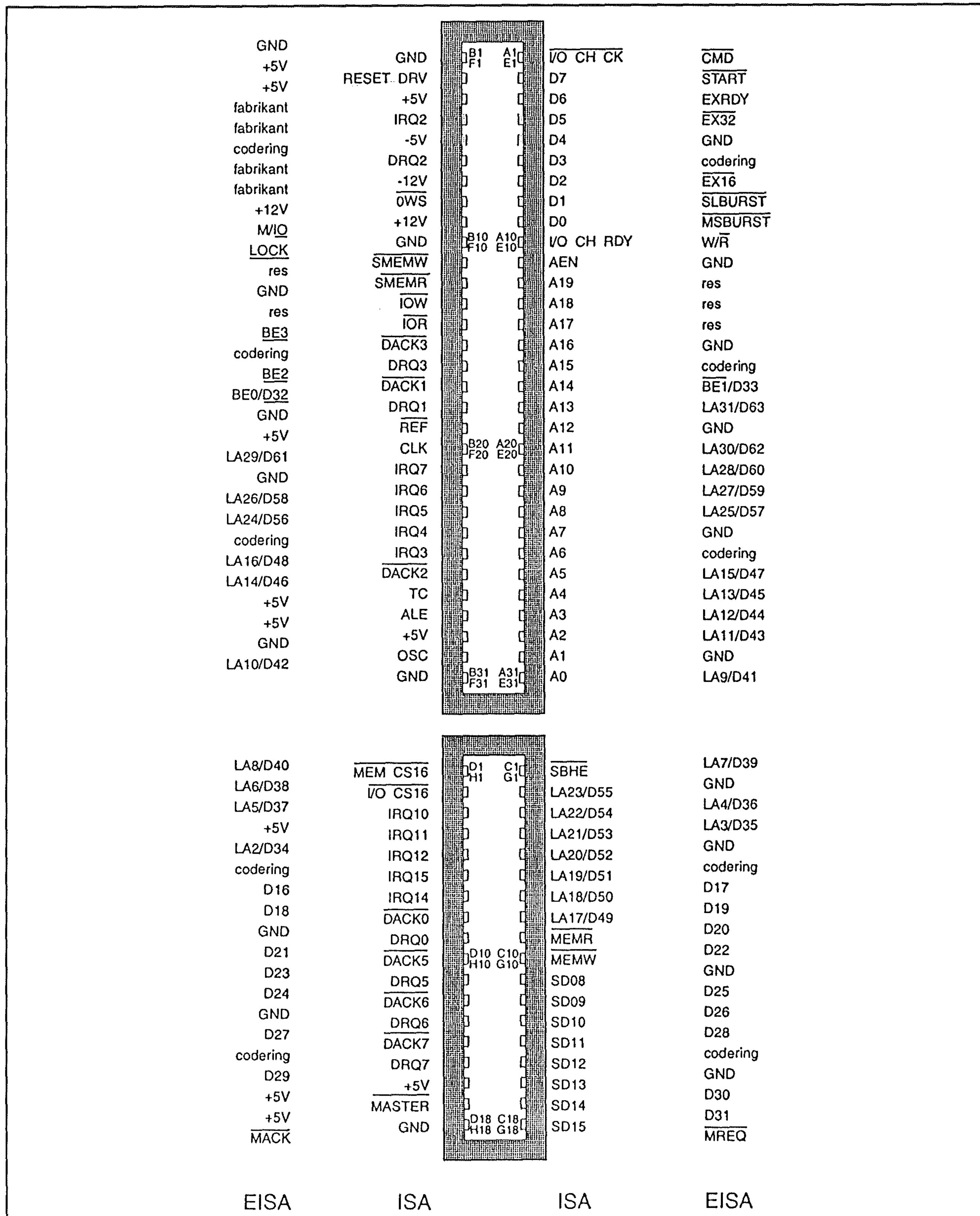
I/O-poorten

Het benaderen van perifere schakelingen via adrespoorten is bij de EISA-architectuur gehandhaafd. Groot verschil met de ISA-architectuur is dat de individuele slot's afzonderlijk benaderd kunnen worden. Dit is noodzakelijk omdat het de bedoeling was de uitbreidingskaarten zonder het instellen van DIP-schakelaars in het systeem te kunnen aanbrengen. Het MSB van het I/O-adres wordt door de logica op het moederbord gedecodeerd, waardoor het juiste slot wordt aangesproken. De adresdecoder op een uitbreidingskaart moet dan alleen nog het LSB decoderen. Op deze manier is het mogelijk dat oude ISA-kaarten zonder problemen in een slot kunnen worden gedecodeerd. In de tabel van figuur 3/19.1-16 is een overzicht gegeven van de toewijzing van de beschikbare I/O-adressen.

Het EISA-slot

Het EISA-slot, zie figuur 3/19.1-17, is een glansstuk der techniek. Het heeft dezelfde grootte als de AT-connector in het ISA-systeem.

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-18: De volledige beschrijving van alle pennen van de EISA-connector.

19.1 Architectuur van de computer hardware en de bus

Hoewel de EISA-connector dezelfde breedte en lengte bezit en alleen enkele millimeter hoger is dan de ISA-connector, zijn alle noodzakelijke leidingen voor de stroomverzorging en aarde alsmede alle besturingslijnen voor ISA- en/of EISA-uitbreidingskaarten van elk formaat aanwezig. Mogelijk werd dit kunststukje door in de connector contacten op twee niveaus aan te brengen. De EISA-contacten liggen daarbij onder de ISA-contacten. Mechanische beschermingen voorkomen dat 8 of 16 bit ISA-kaarten tot de dieper gelegen EISA-contacten kunnen doordringen. EISA-kaarten bezitten inkepingen en kunnen zodoende, aan de beschermingen voorbij, dieper in de connector worden geschoven.
Figuur 3/19.1-18 geeft een volledig overzicht van de functie van alle pennen van het EISA-slot. Iedere penpositie heeft nu uiteraard twee betekenissen, de binnenste benaming geldt voor ISA, de buitenste voor het dieper gelegen EISA-contact.

Local Bus architecturen

Inleiding

MCA- noch EISA-architectuur hebben het doorslaggevende succes gekend dat hun ontwerpers hadden voorspeld. Naast de hoge prijzen is zeker ook de beperkte klokfrequentie van het bussysteem daarvoor verantwoordelijk. Vooral door de algemene acceptatie van Windows werd deze lage overdrachtsnelheid van de bus een steeds groter probleem. Grafisch georiënteerde besturingssystemen, zoals Windows, hebben een zeer grote en regelmatige communicatie tussen het systeem en de beeldscherm-adaptor tot gevolg. Er moet immers heel wat pixel-informatie naar het beeldscherm worden overgebracht als men een venster minimaliseert en een ander venster opent. Omdat zowel ISA, MCA als EISA nog steeds gebruik maakt van een busfrequentie van 10 MHz, kan zelfs de beste grafische kaart de beeldopbouw niet snel genoeg tot stand brengen. Deze lage kloksnelheid van de bussen was een van de voornaamste redenen dat fabrikanten van grafische kaarten hun produkten begonnen te voorzien van intelligente beeldprocessoren, die niet alle beeldinformatie vanuit het systeem moesten betrekken, maar die in staat waren aan de hand van bepaalde instructies zélf delen van het beeld op te bouwen. Maar dat was natuurlijk een noodsprong, veel beter zou het zijn als er een volledig nieuwe systeemarchitectuur werd ontwikkeld, waarbij de slot's werden aangestuurd met dezelfde klokfrequentie als de processor. Een systeem met een 50 MHz 80468 processor zou dan ook met de uitbreidingskaarten op die snelheid kunnen communiceren. Maar natuurlijk moest op de een of andere manier de neerwaartse compatibiliteit gegarandeerd blijven. Intel en MicroSoft hadden het woord "compatibiliteit" immers in de voorbije jaren verheven tot een soort Goddelijk gebod, waar onder geen enkele voorwaarde van afgeweken mocht worden.

LB

Om deze patstelling te doorbreken, hebben enkele fabrikanten in 1992 weer een nieuwe moederbord architectuur gelanceerd onder de naam "Local Bus", afgekort tot "LB". Een LB-systeem is in principe nog steeds gebaseerd op een gewoon ISA-systeem, dat is uitgebreid met één of meer insteekplaatsen voor uitbreidingskaarten die rechtstreeks op de zeer snelle interne bus zijn aangesloten. Op zo'n in-

terne (lokale) insteekplaats staan niet alleen alle 32 datalijnen ter beschikking, zij wordt ook met dezelfde frequentie "geklokt" als de CPU, bij een 50 MHz systeem dus een met de verhoudingsgewijs hoge bus klokfrequentie van eveneens 50 MHz. Daardoor worden via deze speciale insteekplaatsen tot nu toe niet voor mogelijk gehouden overdrachtsnelheden bereikt die, theoretisch, kunnen oplopen tot ongeveer 100 MB/s. In hardware specifieke tests kwamen dergelijke snelheden tenminste binnen bereik. In de praktijk van alledag zal men zich toch met veel lagere snelheden tevreden moeten stellen. LB is dus geen nieuwe versie of verdere ontwikkeling van de moederbordtechnologie zoals dat bij MCA of EISA het geval was, maar eigenlijk slechts een verbeterde variant van de ISA-techniek. In principe onderscheidt zich een LB-slot alleen door de aanwezigheid van extra lijnen die wel in het bus-systeem voorhanden zijn, maar in het standaard ISA-slot niet worden gebruikt.

Systeemintegratie

Moderne LB-systemen presenteren zich echter als meer dan een simpele busuitbreiding of uitgebreide coprocessor poort. Speciale chipset's, met cache, drivers en buffers, integreren de LB-slot's volledig in het PC-systeem. Makers van dergelijke chipset's zijn onder meer bedrijven als Opti, UMC en Headland.

Moederborden met de LB-technologie worden inmiddels door meer en meer producenten aangeboden. Een van de grootste nadelen was dat er in het begin een strijd losbarstte tussen twee systemen:

- VLB;
- PCI.

Op dit moment kan men stellen dat de PCI-technologie deze strijd gewonnen heeft en tot een nieuwe standaard op het gebied van PC-architectuur is verheven.

Algemeen principe van een LB-architectuur

Het principe van de LB-techniek wordt in de figuren 3/19.1-19 schematisch weergegeven. Het bovenste blokschema geeft de structuur van de tot nu toe behandelde architecturen. De slot's worden via de I/O-interface met het systeem verbonden en de klokfrequentie wordt hier gereduceerd tot 8 MHz. Het onderste schema geldt voor een LB-architectuur. De enige uitbreiding is dat er nu een aantal LB-slot's aanwezig zijn, die rechtstreeks zijn aangesloten op de interne bus van het systeem en dus geklokt worden met dezelfde frequentie als de processor.

Bij het vergelijken van de twee figuren wordt zonder meer duidelijk waarin een LB- in wezen van een ISA-architectuur verschilt. Terwijl bij het ISA-bord bijvoorbeeld een grafische kaart via de langzame, maximaal 16 bit brede AT-bus haar werk kan doen, is de grafische kaart bij het LB-bord direct op het snelle 32 bit brede intern bus-systeem aangesloten. De LB-slot's zijn practisch direct met de CPU verbonden en werken met dezelfde klokfrequentie (synchroon).

Voordelen van een LB-systeem

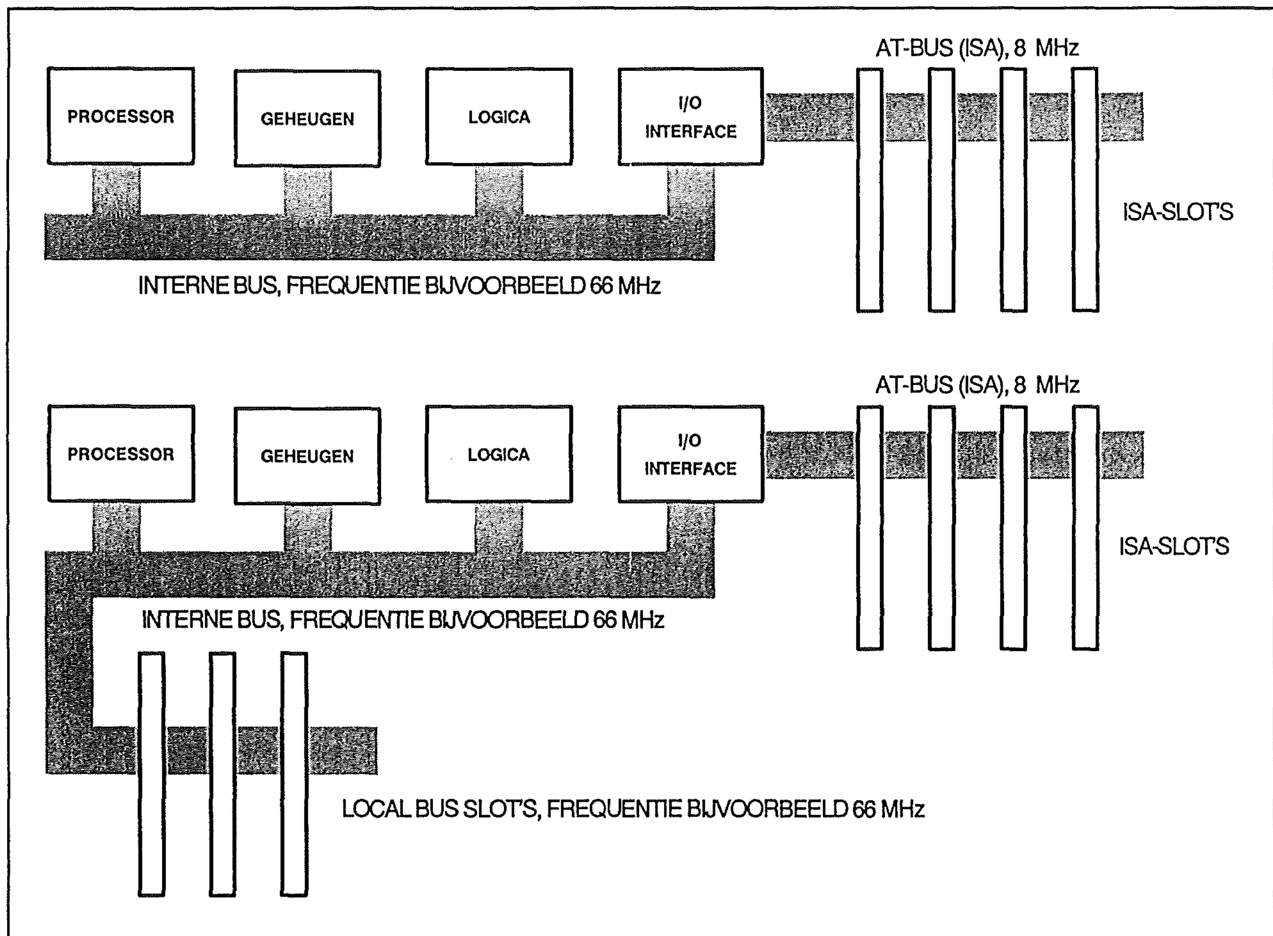
De LB-technologie heeft onderstaande voordelen:

- directe aansluiting op de processor;
- 32 bit gegevensoverdracht;
- hoge klokfrequentie.

Omdat ook een LB moederkaart over normale AT-slot's beschikt, kunnen alle bestaande ISA-kaarten op een moederbord van dit type worden gebruikt. Bij de meeste LB moederborden zijn tegenwoordig drie specifieke LB-slot's aanwezig.

(wordt vervolgd)

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-19: Vergelijking van de tot nu toe behandelde architecturen met deze van een Local Bus.

De VESA Local Bus architectuur

Inleiding

De VESA Local Bus, meestal afgekort tot VLB, is ten opzichte van de later op de markt gebrachte PCI local bus veel meer als een échte lokale bus opgezet. De structuur van VESA sluit direct aan op de lokale processor-bus van een 386, 486 of Pentium. Het gevolg hiervan is bijvoorbeeld dat VESA werkt met buscycli die deze van de 80x86-processoren benaderen. Door de evolutie in de Intel-lijn van processoren kunnen deze buscycli echter nogal verschillen. De 386 kent bijvoorbeeld géén burst-cyclus.

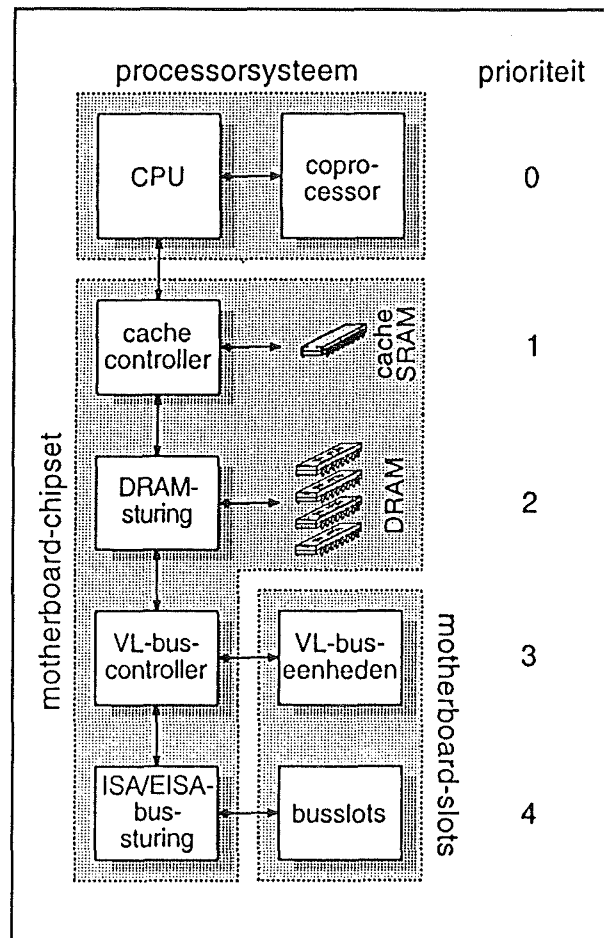
Algemene kenmerken van de VESA bus

De algemene eigenschappen van de VESA locale bus zouden als volgt kernachtig kunnen worden samengevat:

- er bestaat géén ontkoppeling tussen de centrale processor en de uitbreidingsbus;
- het is mogelijk andere bussystemen via een zogenoemde bridge in het systeem te integreren;
- de VLB werkt op de frequentie van de processor: volgens de specificaties betekent dit een maximale busfrequentie van 66 MHz voor eenheden die op het

moederbord geïntegreerd zijn en van 40 MHz (versie 1.0) respectievelijk 50 MHz (versie 2.0) voor eenheden die in een slot zitten;
- VLB heeft standaard een busbreedte van 32 bit met een transferrate van maximaal 133 MB/s (versie 1.0) respectievelijk 160 MB/s (versie 2.0);
- uitbreiding naar 64 bit busbreedte is voorzien dank zij een systeem van multiplexing van signaallijnen: de transferrate stijgt dan tot maximaal 267 MB/s (versie 2.0);
- ook 16 bit brede cycli zijn mogelijk, dank zij de signalen $\overline{LBS16}$ en $\overline{BS16}$;
- VLB ondersteunt burst-transfers tot maximaal vier cycli, waarmee in één cyclus maximaal 32 bytes kunnen worden overgebracht;
- er wordt slechts één host-processor ondersteund;
- VLB werkt met gescheiden adres- en datalijnen voor simultane overdracht van adres- en databytes;
- het systeem ondersteunt gecombineerde VLB/ISA-slots, zodat uitbreidingskaarten in principe ook toegang hebben tot de ISA-signalen;
- VLB ondersteunt dynamische veranderingen van de klokfrequentie;
- er kunnen maximaal drie VL-busmasters naast de VL-buscontroller controle over de bus krijgen;
- ondersteuning van write-back caches.

De algemene opzet van een VLB-systeem
Het blokschema van een VLB-systeem is getekend in figuur 3/19.1-20. Net zoals bij de PCI-bus staat de VL-bus tussen de processor- en het geheugensysteem enerzijds en de standaard uitbreidingsbus anderzijds. Uit het blokschema is af te leiden dat een VL-bus echter niet zo sterk wordt ontkoppeld door het processorsysteem aan de ene en de standaard uitbreidingsbus aan de andere kant. De VL-bus voorziet in maximaal vier VL-bus eenheden, die in de VL-slot's op het moederbord kunnen worden geplaatst.



Figuur 3/19.1-20: Het blokschema van de VLB-architectuur.

Het VLB-systeem wordt door een buscontroller gestuurd. Deze schakeling wekt alle noodzakelijke adres-, data- en stuursignalen op voor de local bus of ontvangt ze van de perifere schakelingen.
De VL-slots hebben een 116-polige connector, waarvan er vier als sleutel worden gebruikt. De VLB-slot's worden steeds 5 mm achter een slot van de stan-

19.1 Architectuur van de computer hardware en de bus

daard uitbreidingbus geplaatst. Een ISA- en VLB-slot zitten dus uitgelijnd achter elkaar, zodat de VLB-kaart niet alleen de signalen van de VL-bus kan verwerken, maar ook toegang heeft tot de signalen van het ISA-slot. Net zoals bij de PCI-bus is de standaard busbreedte 32 bit, die echter ook tot 16 bit kan worden gehalveerd. Daarvoor moet de geadresseerde VL-eenheid het signaal $\overline{\text{LBS16}}$ (Local Bus Size 16) laag maken. Versie 2.0 van de LSB-standaard voorziet bovendien in een uitbreiding van de VL-bus naar een breedte van 64 bit. Hiervoor zijn 37 extra lijnen nodig, die over reeds beschikbare contacten kunnen worden overgedragen via multiplexing. Een lijn voert dan afwisselend twee signalen. Een VLB-eenheid kan een 64 bit cyclus aanvragen met het signaal $\overline{\text{LBS64}}$ (Local Bus Size 64).

Busfrequentie

De VL-bus werkt in principe op de clockfrequentie van de centrale processor. Vanwege de grote verschillen in clocksnelheden zal de VLB-bus dus op een systeemgebonden frequentie werken. Er zijn echter maximale grenzen gesteld aan de busclock. De VL-bus is zo ontworpen dat zij op maximaal 66 MHz kan werken. Dat geldt echter alleen als er geen VL-bus slot's aanwezig zijn en alle aanwezige VL-eenheden op het moederbord zijn ingebouwd. Dempingen, signaalreflecties en paracitaire capaciteiten van de busslot's verhinderen een dergelijk hoge frequentie op de bus zélf. Met uitbreidingsslot's werkt de VL-bus daarom op slechts maximaal 50 MHz, nog steeds het zesvoudige van de EISA busfrequentie! Bij moderne processoren met zeer hoge clockfrequenties zorgt een frequentiedeler ervoor dat de clock-frequentie van de bus gereduceerd wordt tot een veilige waarde.

Write-posting en burst-modus

De geïntegreerde busbuffers van de VL-bus maken het mogelijk de schrijftoegang van de VL-eenheden tot 33 MHz zonder wachtcycli uit te voeren. De VLB heeft dus ook de mogelijkheid tot write-posting. Tijdens het lezen bieden deze buffers geen voordeel. De leestoegang worden daarom altijd met een wait-state uitgevoerd. De VLB-specificatie beschikt ook over een burst-modus, vergelijkbaar met die van de 486 en implementeert de noodzakelijke stuursignalen $\overline{\text{BRDY}}$ en $\overline{\text{BLAST}}$.
Met versie 2.0 kunnen ook continue schrijfacties door een burst uitgevoerd worden. De theoretische maximale transferrate bij een databus-breedte van 32 bit en 50 MHz clock-frequentie (het maximum voor VL-slots) bedraagt tijdens schrijven dus 160 MB/s. Tijdens het lezen is voor de eerste transfer een extra wachtcyclus noodzakelijk, de transferrate daalt dan tot 133 MB/s. De uitbreiding naar 64 bit levert een bijna twee keer zo hoge transferrate op, dus 267 MB/s voor schrijf- en 222 MB/s voor leesbursts.

Wachtcycli

VLB-eenheden kunnen extra wachtcycli aanvragen, bijvoorbeeld als op een bepaald moment geen toegang tot het video-RAM mogelijk is, omdat op de adapter een memory-refresh plaatsvindt of de elektronenstraal geen terugslag uitvoert. De VL-bus beschikt hiervoor over de signalen $\overline{\text{LRDY}}$ (Local Bus Ready) en $\overline{\text{BRDY}}$ (Burst Ready).

Masters en targets

De drie mogelijke VLB-eenheden zijn onderverdeeld in de zogenoemde local bus master (LBM) en local bus targets (LBT). Een LBM kan zelfstandig de sturing van de volledige bus overnemen van de cen-

trale processor of de DMA en datatransfer initialiseren. Een LBT kan dat niet. Een dergelijke eenheid wordt in het verloop van de datatransfer via de VL-bus alleen maar aangestuurd, maar wekt geen eigen bussignalen (behalve LRDY) op. De controle over de busoverdracht wordt door de VLB-controller uitgevoerd, die op het moederbord is geïntegreerd en een arbitrage uitvoert die vergelijkbaar is met deze bij EISA of Microchannel.

Prioriteiten
Aan de verschillende eenheden op het moederbord en in de slot's worden verschillende prioriteiten toegekend. Als een eenheid met een hogere prioriteit de controle heeft, dan kan een eenheid met een lagere prioriteit de besturing van de bus niet overnemen. Maximaal zijn naast de VLB-controller drie VLB-masters beschikbaar.

Acht buscycli
Als een VLB-eenheid wordt geadresseerd om een buscyclus uit te voeren door de processor of een VLB-master dan heeft de geadresseerde eenheid maximaal 20 ns de tijd om toegang te krijgen door het signaal $\overline{LDEVE}$ (Local Device) te activeren. Dat is het teken voor de VLB-controller om de opgevraagde buscyclus uit te voeren. In totaal zijn acht manieren om toegang te verkrijgen mogelijk, dus om een cyclus uit te voeren, die door de stuursignalen M/$\overline{IO}$ (Memory/IO), D/$\overline{C}$ (Data/Command) en W/$\overline{R}$ (Write/Read) worden gecodeerd.
De acht toelaatbare cycli zijn:
- "L-L-L":
  INTA-sequentie;
- "H-L-L":
  instructie-fetching;
- "L-L-H":
  halt/speciale cyclus voor 486;
- "H-L-H":
  halt/shutdown voor 386;
- "L-H-L":
  I/O-leestoegang:
- "H-H-L":
  geheugen-leestoegang;
- "H-H-L":
  I/O-schrijftoegang;
- "H-H-H":
  geheugen-schrijftoegang.

De VLB-cycli voor een enkele transfer en voor INTA-sequenties volgen exact dezelfde regels en fasen als deze bij de 386- en 486-processoren.

Burst-cycli
Zoals reeds geschreven, beschikt VLB over de mogelijkheid om burst-cycli uit te voeren. Hierbij wordt gebruik gemaakt van één adresserings-fase, waarna maximaal 16 bytes, dus vier dubbele woorden kunnen worden overgedragen.
Iedere burst-cyclus wordt door een adresseringsfase ingeleid, waarbij de adreslijnen A2 tot en met A31 een adres overdragen en de $\overline{BEx}$-lijnen alle een signaal met laag niveau voeren. $\overline{BLAST}$ wordt op een hoog niveau gehouden, tot de laatste datatransfer is begonnen. De VL-bus gebruikt een burst-cyclus die overeenkomt met de 486 leesburst. Dit is de enige mogelijke burst-cyclus voor VLB-eenheden van versie 1.0.
De geadresseerde eenheid geeft voor iedere datatransfer een actief $\overline{BRDY}$-signaal terug. Op de ene adresseringsfase volgen een of meerdere datafasen. Hierbij voeren de datalijnen Dx de gegevens en de $\overline{BEx}$-aansluitingen de byte-enable signalen overdragen. Een aktief $\overline{BRDY}$-signaal geeft tijdens de leestoegang aan dat de target-eenheid de gegevens heeft overgedragen.

In versie 2.0 van de VESA-standaard wordt ook een burst-schrijftransfer gespecificeerd. Die werkt op dezelfde manier als een burst-leestransfer, maar dan natuurlijk in omgekeerde richting. De initiator geeft naast de adressen op A2 tot en met A31 ook de gegevens door via D0 ot en met D31. Door gebruik te maken van de busbuffer van de VL-bus kan een write-posting worden uitgevoerd.
Als een VLB-master burst-cycli wil verhinderen, dan moet deze direct na de activering van $\overline{ADS}$ ook het signaal $\overline{BLAST}$ op een laag niveau zetten. De target-eenheid voert dan een enkele transfer uit.

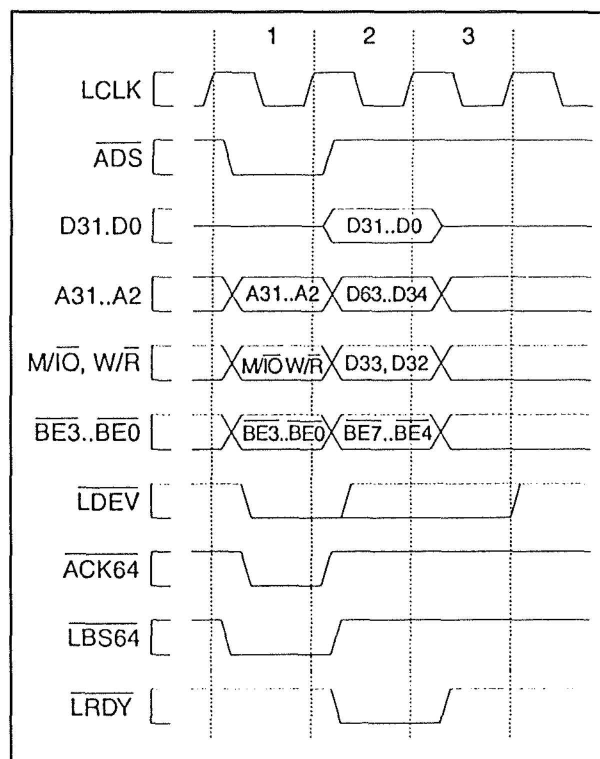
16 bit transfers

In de VLB-slot's zijn de adreslijnen A2 tot en met A31 en de byte-enable signalen $\overline{BE0}$ tot en met $\overline{BE3}$ beschikbaar.
Bij de processoren van de 386SX-reeks beschikt men over de signalen $\overline{BLE}$ (Bus Low Enable) en $\overline{BHE}$ (Bus High Enable) en over het adresbusbit A1. Om de byte-enable signalen $\overline{BE0}$ tot en met $\overline{BE3}$ in de VLB-slot's uit de genoemde processorsignalen af te leiden, moet de VLB-controller voor een 386SX-systeem de volgende logische vergelijkingen uitvoeren:

- $\overline{BE0}$ = [A1 OF $\overline{BLE}$]
- $\overline{BE1}$ = [A1 OF $\overline{BHE}$]
- $\overline{BE2}$ = [$\overline{A1}$ OF $\overline{BLE}$]
- $\overline{BE3}$ = [$\overline{A1}$ OF $\overline{BHE}$]

64 bit transfers

In versie 2.0 is een uitbreiding van de VL-databus naar 64 bit opgenomen, en wel zonder uitbreiding van het VLB-slot of een verhoging van het aantal pennen van de connector.
Deze verbreding van de bus is mogelijk gemaakt door middel van multiplexen van de noodzakelijke extra data- en besturingssignalen op reeds beschikbare lijnen.



Figuur 3/19.1-21: Een voorbeeld van een 64 bit transfer, in dit geval een schrijfactie.

Deze lijnen voeren dus afwisselend twee signalen. De 64 bit transfer is voornamelijk van belang bij burst-cycli. Het voortdurend multiplexen en de daaraan verbonden verlenging van de cyclus bij een enkele transfer zou de winst van de bredere bus immers grotendeels teniet doen. Het multiplexing-principe en het verloop van de 64 bit transfer wordt behandeld aan de hand van een voorbeeld.

64 bit schrijftransfer

In figuur 3/19.1-21 is een enkele 64 bit schrijftransfer voorgesteld. Eerst worden de adressen en stuursignalen op de bus geplaatst en de VLB-controller activeert het adres-strobe signaal $\overline{ADS}$. Bovendien geeft de controller een signaal $\overline{LBS64}$ met een actief laag niveau om de target-eenheid te melden dat een 64 bit transfer

moet worden uitgevoerd. De geadresseerde target accepteert de aanvraag en reageert met een LDEV-signaal om de VL-buscontroller duidelijk te maken dat hij zich als target van de cyclus heeft herkend. Bovendien stuurt de target-eenheid een actief ACK64-signaal terug als hij in staat is de 64 bit transfer uit te voeren. Aan het begin van de tweede LCLK-cyclus worden de te schrijven gegevens klaar gezet. Daartoe plaatst de VLB-controller het laagwaardige dubbele woord op de databus. Bovendien worden A2 tot en met A31, M/IO, W/R en BE0 tot en met BE3 ongeldig gemaakt. In plaats daarvan worden de hoogwaardige dubbele datawoorden D32 tot en met D63 en de bijbehorende byte-enable signalen BE4 tot en met BE7 uitgevoerd. De target neemt de 64 bit over en reageert met een LRDY en een RDYRTN om de overdracht van de enkele cyclus af te sluiten. Het gehele proces heeft twee LCLK-cycli geduurd en is een voorbeeld van de snelst mogelijke enkele transfer van de VL-bus. Als 64 bit in twee gescheiden 32 bit cycli worden verstuurd, dan zijn hiervoor vier LCLK-cycli nodig. Het gevolg is dat in de 64 bit modus de transferrate voor een enkele transfer verdubbelt.

Write-back caches

In specificatie 1.0 is enkel ondersteuning voor write-through caches opgenomen. Het is dan noodzakelijk het Snoop signaal LEADS (Local External Adress Strobe) te activeren als een lees of I/O-cyclus moet worden uitgevoerd. Daardoor wordt automatisch een snoop-cyclus voor alle behalve de juist actieve busmaster gemaakt.

In specificatie 2.0 is echter ook ondersteuning van write-back caches gedefinieerd. Er moet dus ook bij schrijftoegang een toegangscyclus worden uitgevoerd, echter niet voor I/O-cycli omdat anders mogelijk een onnodige write-back wordt uitgevoerd.

Busarbitrage

De VL-bus ondersteunt maximaal drie local busmasters (LBM). Het arbitrageprotocol gaat als volgt. De controle over de VL-bus wordt volgens een vastgelegde prioriteit overgedragen, waarvoor de beide arbitragesignalen LREQ<x> en LGNT<x> dienen. Deze signalen zijn per VLB-slot eenmaal beschikbaar. Een LBM met een hogere prioriteit kan een LBM met een lagere prioriteit de bus ontnemen, maar niet omgekeerd. Als een LBM de bussturing overneemt, dan activeert die een LREQ-signaal. Heeft op een bepaald moment geen LBM met een hogere prioriteit de controle, dan reageert de VLB-controller met een LGNT-signaal en draagt de bus daardoor over. Om de bus weer vrij te geven, deactiveert de actieve LBM zijn LREQ-signaal en de bussturing bevestigt dat door het LGNT-signaal op hoog niveau te zetten.

Als een LBM de bussturing wil overnemen terwijl een andere busmaster de controle heeft, dan activeert die eveneens zijn LREQ-signaal. Bepalend is uiteraard of de nieuwe aanvragende LBM een hogere prioriteit heeft dan de actieve LBM. Is dat niet het geval, dan weigert de VLB-controller de overdracht van de bus en houdt het LGNT-signaal inactief. Heeft de nieuw aanvragende LBM wél een hogere prioriteit, dan deactiveert de VLB-controller het betreffende LGNT-signaal om de actieve LBM duidelijk te maken dat deze de controle over de bus moet overdragen. De actieve LBM sluit het momenteel lopende proces af en deactiveert zijn LREQ-signaal. Nadien kan de aanvragende busmaster met hogere prioriteit de VL-bus overnemen. De VLB-controller ac-

tiveert het betreffende $\overline{\text{LREQ}}$-signaal. Als de nieuwe master zijn opdracht heeft uitgevoerd en zijn $\overline{\text{LREQ}}$-signaal heeft gedeactiveerd, kan de VLB-controller de controle over de bus weer aan de eerste LBM terug geven.

De VLB en DMA

De VLB-architectuur heeft géén directe lijnen voor het aanvragen van een DMA (Direct Memory Access). Dit is een eigenschap van beide local bus architecturen, dus VLB en PCI. Het grote verschil met PCI is dat een VLB-eenheid gewoonlijk ook een ISA- of EISA-slot ter beschikking heeft en de betreffende signalen van deze slot's kan betrekken. Een directie toegang tot het geheugen is dus op de oude manier mogelijk. Zo'n actie wordt dan echter via de (E)ISA-bus en niet via de VL-bus afgehandeld. Met moet dan natuurlijk rekening houden met de beperkingen van deze oude slot-structuren, bijvoorbeeld wat betreft maximale transferrate. Natuurlijk kan een in het geheugenadresgebied van de processor of van de DMA-controller liggende VLB-eenheid de target van een DMA-overdracht zijn, die dan rechtstreeks door de DMA-controller van het moederbord of een busmaster op een uitbreidingskaart gestuurd wordt. Dat heeft dan echter natuurlijk niets te maken met de eigenschappen van de VL-bus, maar alles met de adressering van een eenheid in het geheugenadresgebied. Gebruikt een VLB-eenheid alleen de VL-bus om een data-overdracht uit te voeren, dan moet die zich als een VL-busmaster kunnen gedragen om de VL-bus rechtstreeks te besturen. Deze werkwijze staat toe dat een hogere data-transferrate wordt toegepast dan met de klassieke DMA-techniek mogelijk is, omdat de VL-busmaster natuurlijk de volle 32 of zelfs 64 bit kan gebruiken en de maximale snelheid van de VL-bus kan toepassen. Bij de VLB-architectuur is het traditionele concept van directe geheugentoegang dus door een veel flexibeler en krachtiger systeem, namelijk dat van de externe busmaster, vervangen.

VLB en interrupts

De VLB-specificaties beschikt slechts over één interrupt-lijn, namelijk $\overline{\text{IRQ9}}$. Die werkt niveau-getriggerd met een actief hoog signaal en is direct met de IRQ9 van de (E)ISA-bus verbonden. Normaal gesproken gebruiken de VLB-eenheden voor het aanvragen van een interrupt ook het daarvoor liggende ISA-, EISA- of MCA-slot, zodat genoeg interrupt-lijnen ter beschikking staan. Het is natuurlijk denkbaar dat een fabrikant een pure VLB-adapter zonder contacten voor de (E)ISA-bus of MCA levert. Om ook een dergelijke kaart te mogelijkheid te geven een interrupt aan te vragen is de lijn $\overline{\text{IRQ9}}$ in het VLB-systeem opgenomen.

I/O adresgebied

De VL-bus ondersteunt als lokale bus voor 386-, 486- en Pentium-processoren volgens de specificatie het normale, 64 k groot, adresgebied met 8 bit, 16 bit en 32 bit poorten. Omdat een VL-bus in het algemeen een uitbreiding is voor het ISA- of MCA-systemen, zijn de betreffende poorten ook met de betreffende registers bezet. Alle geldige I/O-adressen en de bijbehorende bezettingen van ISA, EISA en MCA zijn dus bij VLB nog steeds van toepassing. De VLB-specificaties voorzien niet in een vast omschreven configuratieregister op duidelijk vastgelegde I/O-adressen, zoals PCI dat wél heeft. Meestal zijn die adressen fabrikant-afhankelijk beschikbaar, zodat bijvoorbeeld tijdens de

19.1 Architectuur van de computer hardware en de bus

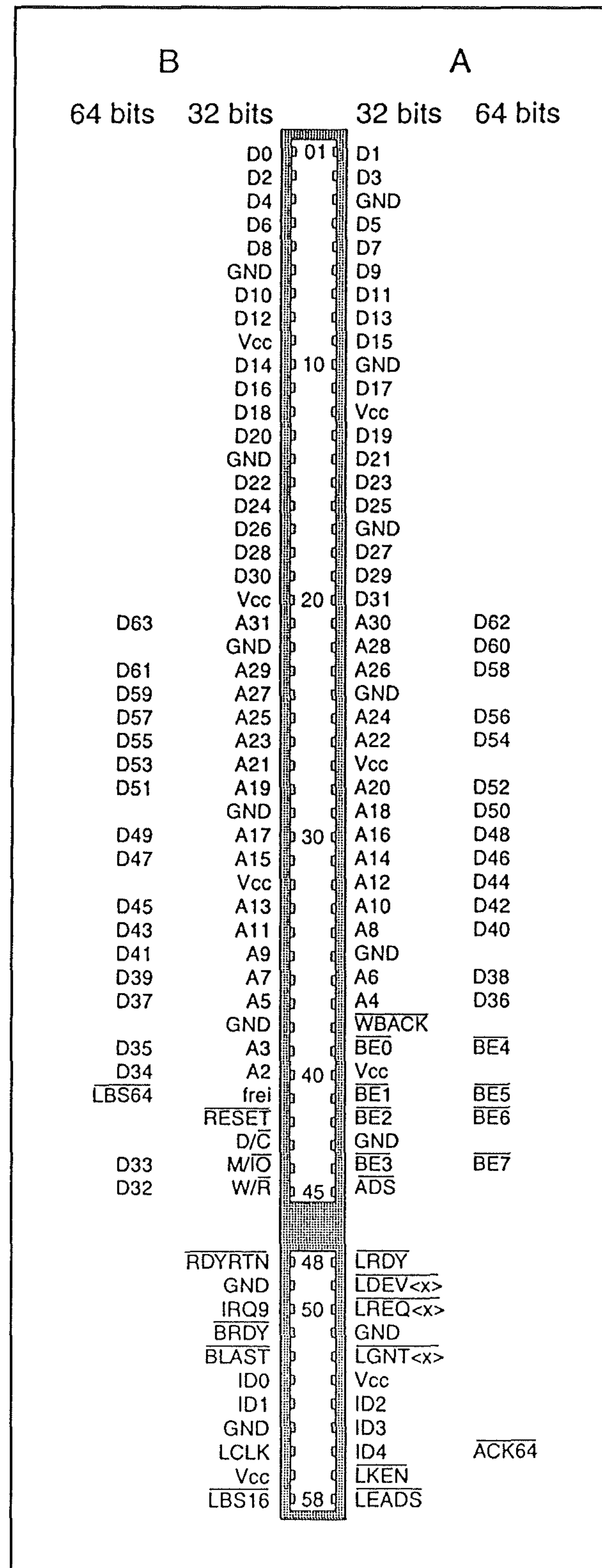
setup het aantal VLB wait-states of de VLB-klokfrequentie kunnen worden ingevoerd.

Het VLB-slot

VLB-slot's beschikken over een 116-polige standaard Microchannel-connector. Ze liggen altijd achter een slot van de standaard uitbreidingsbus, zodat een VLB-kaart altijd signalen en aansluitingen van de standaard slot's kan verwerken (en dat normaal gesproken ook doet om toegang tot de hardware interrupts te verkrijgen). Bij de specificatie 2.0 met de uitbreiding van de VL-bus naar 64 bit worden de daarvoor noodzakelijke nieuwe 37 signalen via de reeds beschikbare aansluitingen met adressignalen gemultiplext. Het aantal VLB-slot's is door hun capacitieve belasting begrensd. Daarbij is het echter niet direct noodzakelijk dat een VLB-systeem ook werkelijk VLB-slot's heeft. De VLB-eenheden, dat zullen dan voornamelijk video-kaarten zijn, kunnen ook op het moederbord geïntegreerd zijn. De gebruiker/ster merkt niets van de interne VLB-architectuur van het systeem, omdat alleen ISA-slot's voor de uitbreidingskaarten ter beschikking staan. Natuurlijk zal een dergelijk systeem een grote flexibiliteit tegenwerken. Vandaar is het gebruikelijk dat VLB-moederborden toch maximaal vier VLB-slot's aan de gebruiker/ster ter beschikking stelt. In figuur 3/19.1-22 wordt een overzicht gegeven van alle contacten van het VLB-slot. De pennen waar twee signaalnamen bij staan, zijn de gemultiplexte pennen die worden gebruikt bij uitbreiding naar 64 bit breedte.

VLB-kaarten

De meeste VLB-kaarten worden vooral als uitbreiding van ISA-moederborden toegepast.



Figuur 3/19.1-22: De aansluitgegevens van het VLB-slot.

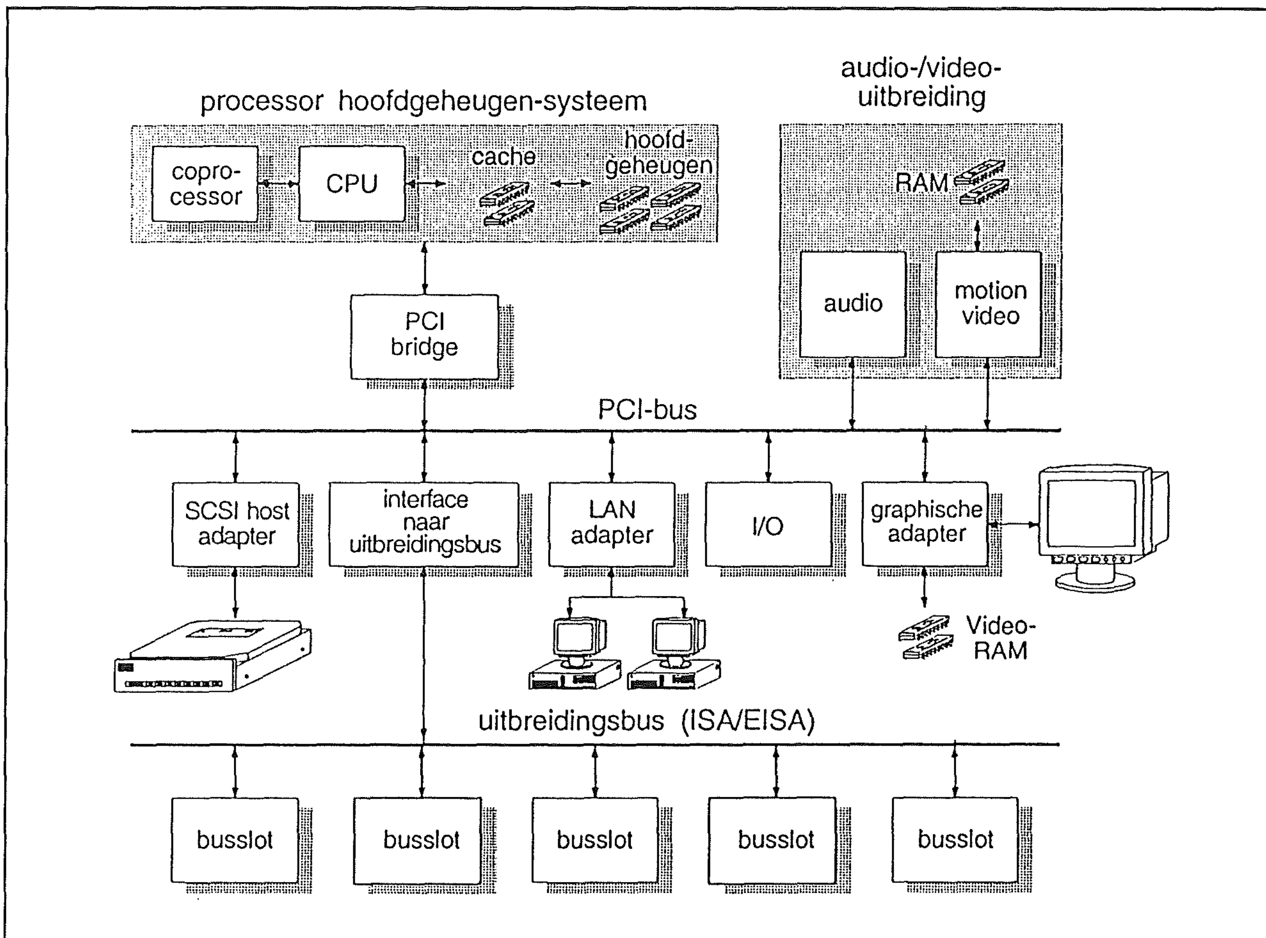
Daardoor bestaat de mogelijkheid oude ISA-kaarten in het systeem te blijven gebruiken. De 16 bit brede databus van deze architectuur is voor de meeste toepassingen (bijvoorbeeld seriële of parallelle kaarten, disketteloopwerk, enzovoort) ruim voldoende.
Alleen die eenheden die een bijzonder hoge overdrachtsnelheid hebben (bijvoorbeeld grafische kaarten en harde schijf controllers) worden dan via de snellere VL-bus gestuurd. Daardoor verkrijgt men op een uiterst goedkope manier een flexibel (ISA) en bovendien snel (VLB) systeem.

De PCI-architectuur

Inleiding
PCI is het letterwoord voor "Peripheral Component Interconnect". Deze local bus werd ontwikkeld door Intel, met uiteraard als groot voordeel dat deze belangrijkste IC-producent van het PC-wereldje onmiddellijk ondersteunende chip-set's kon aanbieden.
De algemene kenmerken van een PCI-systeem kunnen als volgt worden samengevat:
- ontkoppeling van de processor en de PCI-bus door het toepassen van een zogenoemde "bridge";
- de standaard 32 bit busbreedte geeft een theoretisch maximale transfer-rate van 133 MB/s;
- de uitbreiding tot 64 bit busbreedte geeft een theoretisch maximale transfer-rate van 266 MB/s;
- zogenoemde "burst-mode" voor zeer snelle data-overdracht;
- het systeem ondersteunt automatisch uitbreidingskaarten die werken op voedingsspanningen van +5 V of van +3,3 V;
- het systeem geeft de mogelijkheid dat kaarten de controle over de bus overnemen, het zogenoemde "busmastering"-principe;
- zowel de adres- als de data-bus worden gemultiplexed, hetgeen betekent dat één aansluiting van de bus zowel een adresserings- als een gegevens-functie heeft, met als voordeel een zeer efficiënt gebruik van het aantal aansluitingen;
- de PCI-bus ondersteunt zowel de oude ISA-, EISA- als MCA-kaarten;
- de volledige bus kan geconfigureerd worden via software, die de inhoud van registers beïnvloedt;
- de PCI-bus is processor-onafhankelijk.

De structuur van de PCI-bus
De algemene structuur van een PCI-bus is getekend in figuur 3/19.1-23. Het belangrijkste kenmerk van de PCI-structuur is de absolute scheiding tussen enerzijds het deel processor/geheugen en anderzijds de perifere schakelingen. Verantwoordelijk voor deze scheiding is de "PCI-bridge". Deze brug kan opgevat worden als een transparante interface tussen twee bussystemen.
Transparant wil zeggen dat de scheiding alleen hardwarematig bestaat.
Softwarematig kan de scheiding via de transparante brug overwonnen worden. De ontkoppeling van processor en rest van het systeem gaat zo ver, dat processor en brug parallel kunnen werken. Zo is het bijvoorbeeld mogelijk dat, als de processor even geen PCI-eenheden moet benaderen maar een geheugencyclus uitvoert, de PCI-bridge volledig zelfstandig gegevens tussen twee PCI-eenheden organiseert.

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-23: Het blokschema van een PC met PCI-structuur.

Aan de PCI-bus worden "PCI-agents" gekoppeld: perifere apparatuur die rechtstreeks gebruik wenst te maken van de eigenschappen van PCI, zoals de hoge transfer-rate en de mogelijkheid tot busmastering. Typische voorbeelden van PCI-agents zijn SCSI-adapters, LAN-adapters, grafische kaarten en de I/O-adapter. Tegenwoordig zijn de meeste PCI-agents in het moederbord geïntegreerd. Het gevolg is dat een typisch PCI-moederbord slechts drie echte PCI-slot's heeft, waarop men bijvoorbeeld een snelle data acquisitie kaart of een multi-media kaart met live video kan aansluiten. Dergelijke kaarten vergen veel rekencapaciteit en snelle transfer, omdat het op het scherm zetten van live video betekent dat minstens 25 keer per seconde een volledig beeld moet worden gedecomprimeerd. Naast de snelle PCI-bus heeft ieder PCI-systeem ook nog een paar "oude" slot's. Deze slot's kunnen zowel werken volgens de ISA-, EISA- of MCA-architectuur. Deze derde, trage bus wordt via een speciale interface met de echte PCI-bus verbonden. Deze interface wordt door de PCI-architectuur gezien als een gewone PCI-eenheid.

Multiplexing

Een tweede eigenschap van de PCI-architectuur is een ver doorgedreven toepassing van multiplexing. Bij multi-

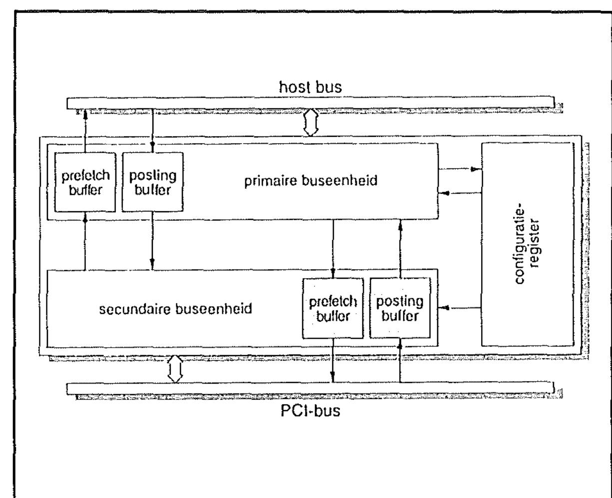
19.1 Architectuur van de computer hardware en de bus

plexing worden twee of meerdere signalen verzonden via één lijn. Bij de PCI-structuur worden buslijnen afwisselend gebruikt voor het versturen van adres- en van data-gegevens. Dit heeft als groot voordeel dat de bus heel wat minder lijnen nodig heeft dan zonder multiplexing het geval zou zijn. Nadeel is uiteraard dat het verzenden van gegevens of adressen langer duurt. Er zijn nu twee of drie klokcycli nodig voor het verzenden van gegevens. Tijdens de eerste cyclus wordt het adres overgebracht en tijdens de tweede en/of derde de gegevens. De maximale praktische transfer-rate bedraagt hierdoor slechts 66 MB/s tot 44 MB/s.

Burst mode

Door het invoeren van het multiplex-principe hebben de ontwerpers van het PCI-systeem de theoretisch haalbare maximale transfer-rate bewust terug geschroefd. Om toch op de maximale snelheid te kunnen werken heeft men de "burst mode" ingevoerd. Bij deze mode wordt het adres in één cyclus overgebracht. Nadien zorgen speciale schakelingen in de zender en de ontvanger ervoor dat het adres bij iedere klokcyclus automatisch met een wordt opgehoogd. Het gevolg is dat het momentele adres altijd bekend is. Nadien kunnen de gegevens in diverse opeenvolgende cycli worden overgebracht. De burst mode heeft geen beperkingen wat betreft het aantal overdrachten dat wordt uitgevoerd. Op deze manier stijgt de maximale transfer-rate tot 133 MB/s bij 32 bit overdracht of tot 266 MB/s bij 64 bit overdracht.

Een belangrijke eigenschap van de PCI-bus is dat burst-cycli door de PCI-bridge volledig zelfstandig worden samengesteld. Zoals uit figuur 3/19.1-24 volgt, bevat de PCI-bridge intelligente lees- en schrijfbuffers. De brug kan enkelvoudige schrijf- en leescycli combineren tot burstcycli, onder de voorwaarde dat de adressen van de acties op elkaar volgen. Dit is een zeer handig systeem om bijvoorbeeld zo snel mogelijk grote hoeveelheden data over te brengen naar de grafische kaart, bijvoorbeeld voor het opbouwen van een Windows-venster op het scherm.



Figuur 3/19.1-24: De blokschematische voorstelling van een PCI-bridge.

Buscycli van de PCI-bus

Dank zij de intelligente PCI-bridge kent de PCI-bus veel meer soorten bus-cycli dan de overige bussen. PCI heeft niet minder dan twaalf cycli, die worden gedefinieerd door de signalen $C/\overline{BE0}$ tot en met $C/\overline{BE3}$. Een overzicht van de twaalf verschillende cycli, met hun $C/\overline{BE}$-code:

- "L-L-L-L":
  INTA-cyclus;
- "L-L-L-H":
  speciale cyclus;
- "L-L-H-L":
  I/O-leescyclus;
- "L-L-H-H":
  I/O-schrijfcyclus;
- "L-H-H-L":

19.1 Architectuur van de computer hardware en de bus

leescyclus naar het geheugen;
- "L-H-H-H":
  schrijfcyclus naar het geheugen;
- "H-L-H-L":
  leescyclus in verband met PCI-configuratie;
- "H-L-H-H":
  schijfcyclus in verband met PCI-configuratie;
- "H-H-L-L":
  meervoudige leescyclus van het geheugen;
- "H-H-L-H":
  duale addresseringscyclus;
- "H-H-H-L":
  line geheugencyclus;
- "H-H-H-H":
  geheugen schrijfcyclus met invalidatie.

Cyclus-verloop

Een cyclus zal steeds gegevens transporteren van de ene naar de andere PCI-eenheid. De aanvragende eenheid wordt "initiator" genoemd, de geadresseerde eenheid de "target". Iedere cyclus wordt voorafgegaan door het activeren van het $\overline{\text{FRAME}}$-signaal. De target geeft aan dat hij klaar staat door het actief maken van het $\overline{\text{TRDY}}$-signaal. Bij een schrijfcyclus betekent dit dat de target klaar staat om de gegevens op de lijnen AD0 tot en met AD31 (respectievelijk AD0 tot en met AD63) te ontvangen. Bij een leescyclus betekent dit dat de target de gevraagde gegevens reeds op de reeds genoemde lijnen heeft gezet. Natuurlijk moet ook de initiator een "ik ben klaar"-signaal opwekken. Dat is het $\overline{\text{IRDY}}$-signaal, dat aangeeft dat de initiator de gegevens respectievelijk op de ADx-lijnen heeft gezet dan wel in staat is de gegevens op die lijnen te ontvangen. Nadien volgt een adresfase, waarbij op de adres/data-lijnen (ADx) een adres wordt gezet en de C/$\overline{\text{BE}}$x-lijnen de cycluscode doorgeven. Tot slot volgen een of meerdere data-fasen, waarbij op de ADx-lijnen de gegevens verschijnen en de C/$\overline{\text{BE}}$x- lijnen de enable-signalen transporteren. De cyclus wordt gestopt als de initiator het $\overline{\text{FRAME}}$-signaal weer deactiveert. Maar ook de target is in staat een cyclus af te breken door het $\overline{\text{STOP}}$-signaal te activeren. Dit wordt in het PCI-jargon een "target abort" genoemd.

De INTA-cyclus

De INTA-cyclus is bedoeld voor het rechtstreeks adresseren van een interrupt-controller, dus zonder dat een specifiek adres wordt aangegeven. De controller herkent de cyclus zelfstandig en reageert hier onmiddellijk op.

De speciale cyclus

Deze cyclus wordt gebruikt om algemene codes aan alle PCI-agenten door te geven, bijvoorbeeld over de status van de processor (halt, shutdown). Deze codes zitten vervat in de adreslijnen AD0 tot en met AD15. De volgende codes zijn gespecificeerd:
- 0000_{HEX}:
  processor shutdown
- 0001_{HEX}:
  procesor halt
- 0002_{HEX}:
  specifieke code van de x86-familie

De overige codes van 0003_{HEX} tot en met $FFFF_{HEX}$ zijn gereserveerd. Bij specifieke codes van de x86-familie moet gedacht worden als instructies als INTA, FLUSH, etc.

De I/O leescyclus

Deze cyclus leest gegevens van een PCI-eenheid uit de I/O-adresruimte. Deze adressen verschijnen op de lijnen ADx, waarna de gegevens volgen.

De I/O schrijfcyclus
Deze cyclus schrijft gegevens van een PCI-eenheid naar de I/O-adresruimte. Ook nu wordt er eerst het adres op de ADx-lijnen gezet en volgen nadien de gegevens.

De leescyclus van het geheugen
Deze cyclus adresseert een PCI-eenheid in de geheugen adresruimte en leest nadien de ter beschikking gestelde gegevens. Het adres wordt op de lijnen ADx gedefinieerd als een dubbeladres, de byte-enable signalen C/$\overline{\text{BEx}}$ geven aan welke bytes van de 32 of 64 bit brede databus geldig zijn.

De schrijfcyclus van het geheugen
Ook deze cyclus adresseert een PCI-eenheid in de geheugen adresruimte en schrijft er gegeven naar toe. De functie van de lijnen ADx en C/$\overline{\text{BEx}}$ is identiek als bij de vorige cyclus.

De configuratie leescyclus
Via de lijn IDSEL wordt een PCI-eenheid geselecteerd. Nadien wordt de configuratie adresruimte van deze geselecteerde eenheid aangesproken door middel van de adressignalen AD2 tot en met AD7. Bij multifunctionele eenheden heeft men hierbij de hulp van de signalen AD8, AD9 en AD10, die de te adresseren subeenheid definiëren. Ook nu bestaat een volledige cyclus uit een adres- en een data-fase.

De configuratie schrijfcyclus
Deze werkt op dezelfde manier als een configuratie leescyclus. Maar nu schrijft uiteraard de initiator gegevens in de configuratieruimte van de target.

De line geheugencyclus
Deze cyclus is een uitbreiding op de "gewone" benadering van het geheugen. Er zijn minder transfers nodig om een gelijke hoeveelheid gegevens over te dragen. Bij dit soort cycli zal de initiator meer dan twee 32 bit brede datablokken uit het geheugen willen lezen. Deze blokken worden achter elkaar verzonden, zodat een dergelijke cyclus bestaat uit één adresseerfase, gevolgd door meerdere data-fasen.

De meervoudige leescyclus van het geheugen
Deze cyclus is een uitbreiding van de line cyclus. De initiator geeft nu te kennen dat het de bedoeling is een omvangrijke hoeveelheid gegevens sequentieel (opeenvolgend) uit het geheugen te lezen. Dit komt bijvoorbeeld voor als het de bedoeling is een deel van het video-geheugen op te bergen in het systeemgeheugen. Deze cyclus bestaat dus uit één adresseringsfase, gevolg door een groot aantal datafasen. De cyclus eindigt als de initiator het signaal $\overline{\text{FRAME}}$ de-activeert.

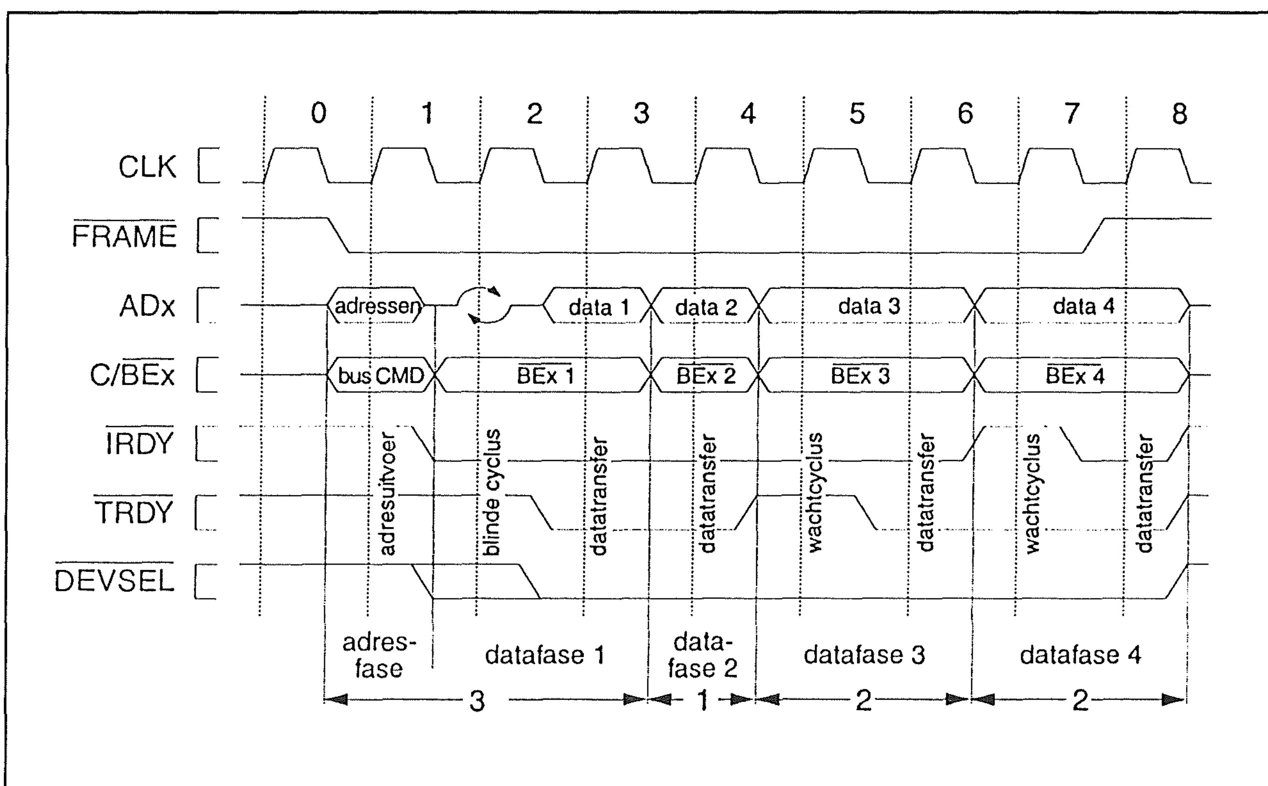
De duale addresseringscyclus
Deze cyclus is speciaal ontwikkeld om een 64 bit breed adres door te geven aan een PCI-eenheid, die over een slechts 32 bit breed adrespad beschikt. Het 64 bit breed adres wordt hiervoor opgesplitst in twee 32 bit delen, het MSB (meest belangrijke deel) en het LSB (minst belangrijke deel).

De geheugen schrijfcyclus met invalidatie
Met deze cyclus geeft de initiator aan, dat minstens een volledige cache-line moet worden doorgegeven. Een cache-line bestaat uit diverse databytes, die in het snelle cache-geheugen zijn opgeborgen. Cachelines zijn meestal 16 of 32 byte groot. Het gevolg van een dergelijke cyclus is dat de inhoud van het systeemgeheugen actueler wordt dan de inhoud van het cache-

geheugen. De betreffende cache-line kan dan geïnvalideerd worden, hetgeen betekent dat men geen write-back hoeft toe te passen. De gegevens in het systeemgeheugen zijn immers geactualiseerd!

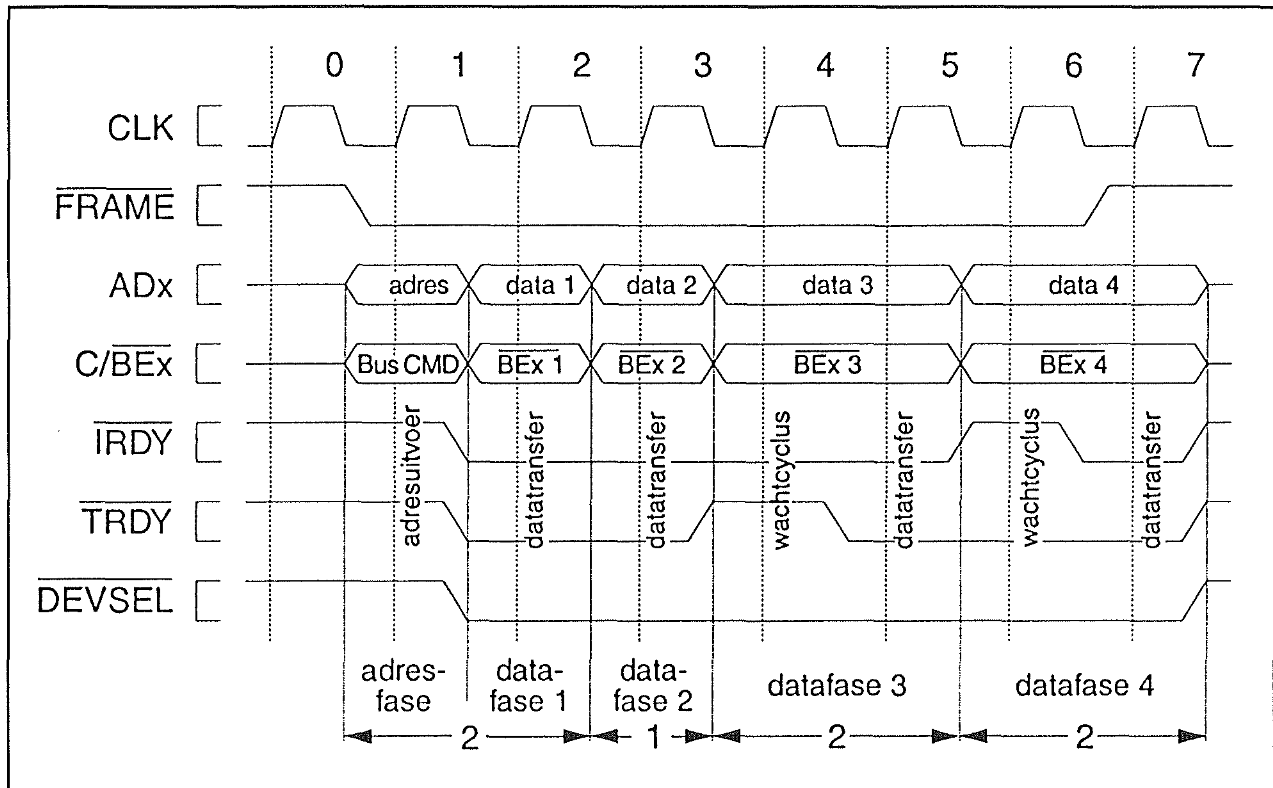
De leescyclus
In figuur 3/19.1-25 is het typische verloop van een PCI leescyclus getekend. De eerste clock-cyclus na de adreseringsfase wordt gebruikt voor het veranderen van de transferrichting op de gemultiplexte bus. Gedurende de adresseringsfase leveren de ADx-lijnen een waarde, namelijk het adres, terwijl op dezelfde lijnen in de datafase een waarde (de gegevens) te vinden is. De eerste waarde kan dus op zijn vroegst in de derde cyclus worden overgenomen.

De schrijfcyclus
Een schrijfcyclus gaat ongeveer op dezelfde manier als een leescyclus. Nu wordt de actie ingeleid door het actief worden van het signaal $\overline{\text{FRAME}}$. Het grote verschil is dat de gemultiplexte adres/data-bus niet omgeschakeld moet worden. Dat betekent dat de gegevensfase onmiddellijk na de adresseringsfase kan volgen. De eerste gegevens kunnen dus reeds bij de tweede klokcyclus worden overgedragen. In de diagrammen van figuur 3/19.1-26 zijn, ter illustratie, twee wachtcycli ingevoerd, die worden veroorzaakt door het inactief blijven van de signalen $\overline{\text{TRDY}}$ en $\overline{\text{IRDY}}$.
De geadresseerde PCI-eenheid reageert met een actief $\overline{\text{DEVSEL}}$-signaal, wanneer de eenheid zich als target van de datatransfer heeft geïdentificeerd.



Figuur 3/19.1-25: De leescyclus van een PCI-bus.

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-26: Het typische verloop van een schrijfcyclus bij de PCI-architectuur.

Bus-arbitrage bij PCI
Bij de eerste bus-systemen kon alleen de processor en de DMA-unit, de direct memory transfer unit, de controle over de bus in handen nemen. Het was dus onmogelijk dat uitbreidingskaarten, buiten de processor om, toegang tot de bus kregen. Bij de MCA- en EISA-bussen werd voor het eerst het principe van "bus-arbitrage" toegepast. Bij dit systeem kunnen perifere processoren de controle over de bus overnemen van de systeemprocessor. Een intelligente kaart kan dus, via de bus, rechtstreeks toegang krijgen tot het systeemgeheugen, zonder dat de systeem-processor daar iets mee te maken heeft. De processor die op een bepaald moment de bus controleert wordt de "bus-master" genoemd. Het zal duidelijk zijn dat dit systeem een heleboel voordelen biedt.

Ook het PCI-systeem beschikt uiteraard over bus-arbitrage. De bus-arbitrage is echter voor iedere cyclus gescheiden, hetgeen tot gevolg heeft dat de bus-master de PCI-bus tussen twee cycli niet onder controle kan houden. Men zou denken dat deze manier van arbitrage het systeem vertraagt. Dit is echter niet het geval, omdat bij PCI wordt gewerkt met "hidden arbitrage". De arbitrage wordt als het ware verstopt achter de cyclus die op een bepaald moment bezig is. Nadat deze cyclus is afgesloten zal de controle over de bus worden overgedragen aan de bus-master die hierom heeft gevraagd. Hiervoor staan twee signalen ter beschikking, namelijk $\overline{REQ}$ en $\overline{GNT}$. Iedere bus-master in het systeem beschikt over eigen $\overline{REQUEST}$- en $\overline{GRANT}$-signalen, die door een centrale arbitrage-logica worden ge-

ëvalueerd. Een bus-master moet het signaal $\overline{\text{REQ}}$ activeren om het overnemen van de bus aan te kondigen. Bovendien moet de arbitrage-logica het signaal $\overline{\text{GNT}}$ activeren voordat een master de controle over de bus toegewezen krijgt.

DMA bij de PCI

DMA is het letterwoord voor "Direct Memory Access", het rechtstreeks benaderen van het geheugen. Dit is een zeer tijdbesparende techniek, omdat het hierdoor mogelijk is grote hoeveelheden gegevens rechtstreeks van de periferie naar het systeemgeheugen over te brengen. EISA en MCA stelden hiervoor speciale signalen ter beschikking, die de directe toegang tot het geheugen vanuit de periferie regelden. PCI doet dit niet, omdat deze directe geheugen toegang in feite niet noodzakelijk is. Het concept van de PCI-bus biedt andere, krachtiger mogelijkheden om vanuit de periferie het geheugen aan te spreken.

De gegevensoverdracht wordt daarbij bijvoorbeeld door het DREQ-signaal getriggerd in de enkele overdrachtmodus. De noodzakelijke busbesturingssignalen $\overline{\text{IOW}}$, $\overline{\text{MEMR}}$ worden door de DMA-controller (de busmaster) opgewekt, zodat de centrale systeemprocessor in deze overdracht geen rol speelt. Men kan het samenwerken van de (busmaster-) DMA-controller met het triggersignaal DREQ als een busmaster-cyclus van de PCI-eenheid opvatten, waarbij de busmasterchip zich standaard op het moederbord bevindt en daar de busbesturingssignalen genereert, terwijl de adapter deze tweede busmaster met behulp van het DREQ-signaal gebruikt. Een eigen busmasterchip op de adapter zelf maakt dit enigszins omslachtige traject overbodig: de adapterbusmaster kan nu alle signalen zelf genereren en daardoor op een willekeurige manier de I/O- en adresruimte adresseren. Dat vraagt echter om busarbitrage die zich ook over de uitbreidingsbus (de slot's) uitstrekt, terwijl bij DMA de arbitrage alleen tussen processor en DMA-controller op het moederbord plaatsvindt. Het gebruik van een externe busmaster is dus flexibeler en krachtiger dan DMA, maar maakt ook een complexe arbitrage noodzakelijk. Verder is het mogelijk dat de externe busmaster op de PCI-kaart een soort DMA-controller voorstelt die snel gegevens tussen geheugen en adapter kan uitwisselen.

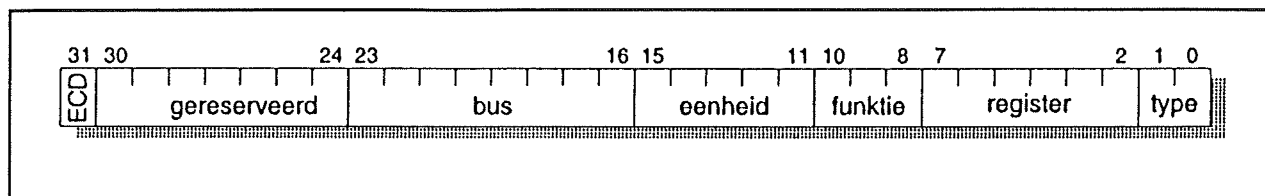
Dit specifieke PCI busmaster-concept is algemener dan de directe geheugenbenadering en is een gestandaardiseerd onderdeel van het PCI-systeem.

Interrupts en PCI

Interrupts zijn onderbrekingen in de normale werking van de systeemprocessor, die kunnen worden aangevraagd door perifere apparatuur. Dank zij een interrupt kan bijvoorbeeld het toetsenbord worden gelezen of de gegevens die door de muis worden geleverd verwerkt. Interrupts vormen een optioneel onderdeel van de PCI-structuur, die laag actief werken. Een PCI-eenheid krijgt de interruptlijn $\overline{\text{INTA}}$ toegewezen. De drie overige interruptlijnen, $\overline{\text{INTB}}$, $\overline{\text{INTC}}$ en $\overline{\text{INTD}}$, kunnen gebruikt worden door multifunctionele PCI-eenheden.

De interrupts worden in de PCI-brug aangepast aan het interrupt-systeem met de IRQx-lijnen van de AT-structuur. Dat gebeurt met behulp van de SETUP in de computer-BIOS. Afhankelijk van het slot waarin bijvoorbeeld een IDE harde schijf controller zit, moet de bijbehorende interrupt $\overline{\text{INTA}}$ voor dit slot worden vertaald naar IRQ14 van de AT-architectuur.

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-27: De samenstelling van het CONFIG_ADDRESS register.

De $\overline{\text{INTA}}$ van een ander slot (bijvoorbeeld een netwerkkaart) kan dan naar een andere AT-IRQ worden geschreven, bijvoorbeeld IRQ11. Welke IRQ de PCI-interrupt activeert moet softwarematig worden vastgelegd. Hierdoor vervalt de strakke vastlegging van IRQ's, zoals dat bij de oude busstructuren het geval was. Bovendien worden lijnen naar de slot's uitgespaard, er zijn nu maar immers vier interruptlijnen noodzakelijk, terwijl de oude bussen ISA, EISA en MCA elf interruptlijnen moesten hebben.

I/O adresruimte en PCI

De PCI-bus ondersteunt zowel 32 als 64 bit I/O adresruimte. Dat geldt alleen voor zuivere PCI-eenheden. Deze eenheden kunnen echter boven 64 k niet meer door een 80x86-processor worden aangesproken, omdat deze processoren slechts een 16 bit I/O adres ondersteunen. De poorten in een systeem met PCI-bus liggen alle onder 64 k, indeling en adressen van de poorten is ten opzichte van EISA of MCA niet veranderd. De oude standaarden ISA, EISA en MCA moeten immers met de PCI-structuur samenwerken.

De PCI-structuur bevat twee registers in een I/O-gebied dat bij EISA voor het moederbord is gereserveerd. Deze twee 32 bit registers CONFIG_ADDRESS en CONFIG_DATA worden gebruikt voor de benadering van het configuratie-adresgebied en liggen op de adressen $OCF8_{hex}$ (ADDRESS) en $OCFC_{hex}$ (DATA).

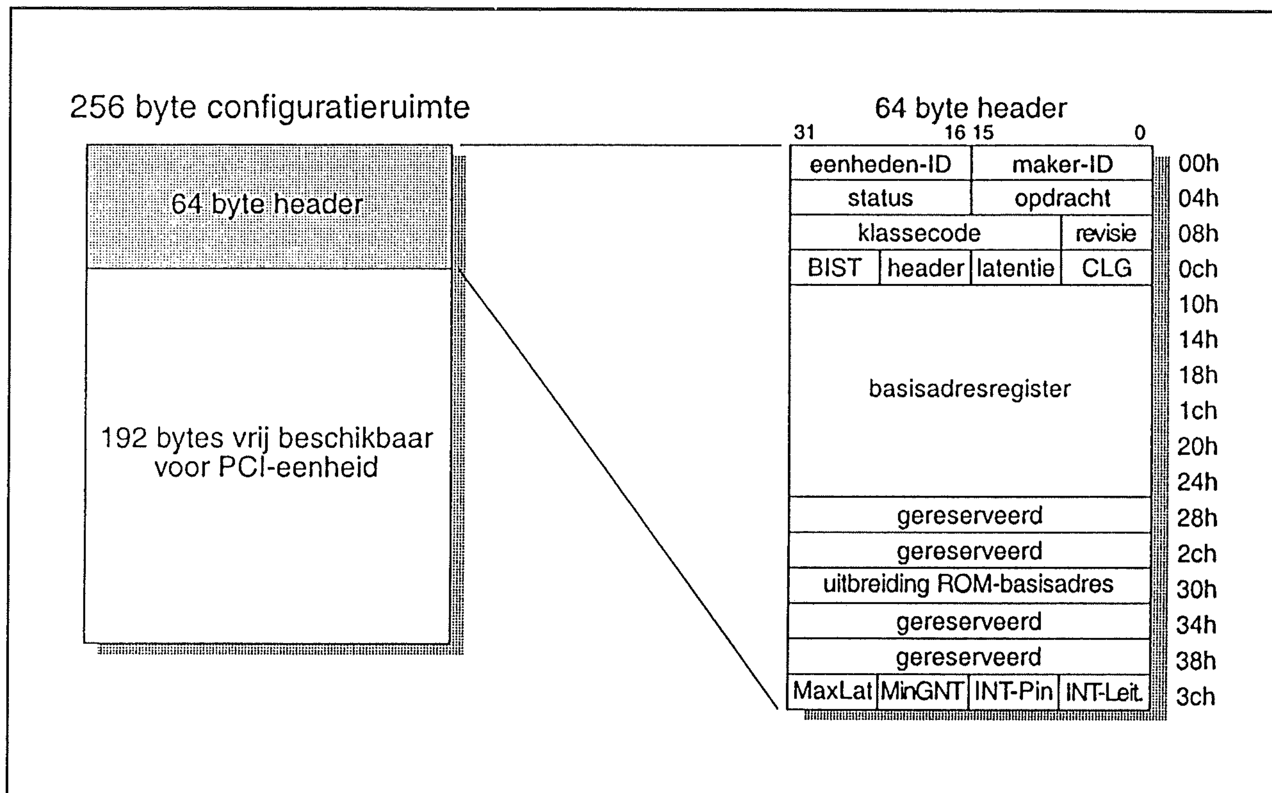
Wanneer het de bedoeling is een dubbelwoord in de configuratieruimte van een PCI-eenheid te lezen of te schrijven moet men eerst het gewenste adres in het CONFIG_ADDRESS-register plaatsen. Het nadien schrijven naar CONFIG_DATA stuurt de waarde door naar de vastgelegde plaats in de configuratie-adresruimte van de geadresseerde eenheid, lezen vraagt de waarde op. De samenstelling van het register CONFIG_ADDRESS is weergegeven in figuur 3/19.1-27.

Wanneer het bit ECD wordt geset, voert de PCI-brug voor het I/O-adres $OCFC_{hex}$ (CONFIG_DATA) een configuratiecyclus uit.

De betekenis van de bits in het register CONFIG_ADDRESS is als volgt:

- Bus:
  Het veld Bus bepaalt het nummer van de PCI-bus in een hiërarchisch gestructureerd systeem met maximaal 256 PCI-bussen. Zo'n uitgebreid systeem kan door diverse PCI-bruggen worden geïmplementeerd.
- Eenheid:
  Eenheid selecteert een van de 32 mogelijke PCI-agenten.
- Functie:
  Functie selecteert een van maximaal acht functies binnen een multifunctionele PCI-eenheid.
- Register:
  Register adresseert een van de 64 mogelijke DWord-velden in de aangegeven configuratie-adresruimte.

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-28: De samenstelling van de configuratieruimte.

- Type:
  Type wordt gebruikt om de brug mee te delen of de geadresseerde eenheid is aangesloten (Type="L-L"), en of er na decodering van de eenheid- en functievelden een configuratie-cyclus moet worden uitgevoerd. Als de inhoud van Type gelijk is aan "L-H", dan bevindt de eenheid zich achter de brug en kopieert de brug de inhoud van CONFIG_ADDRESS onveranderd naar de adres/databus. Een benadering van de configuratie-adresruimte via de twee registers CONFIG_DATA en CONFIG ADDRESS wordt configuratiemechanisme #1 genoemd.
  Men kan de brug daarmee ook opdragen een speciale cyclus uit te voeren, door in het adresregister voor het busveld de PCI-brug te specificeren en de velden voor eenheid en functie met enen en het registerveld met nullen te vullen. De eerstvolgende benadering van het dataregister resulteert dan in een speciale cyclus, waarbij de lijnen AD0 tot en met AD31 de dataregisterwaarde overdragen.

De configuratie adresruimte

Iedere PCI-eenheid en iedere eenheid die binnen een multifunctionele eenheid op zichzelf staat heeft een configuratieruimte van 256 byte, samengesteld uit 64 registers van 32 bit. De samenstelling van deze configuratieruimte is getekend in figuur 3/19.1-28.

Aan het begin staat een header van 64 byte, die voor iedere eenheid dezelfde structuur heeft. De invulling van de overige 192 bytes is afhankelijk van de betreffende PCI-eenheid. Deze wordt bepaald door de configuratiesoftware van de een-

heid. De velden in de configuratieruimte zijn ingedeeld volgens het "little endian"-formaat.

De samenstelling van deze ruimte is als volgt:

- Eenheden-IF:
  Geldige waarden voor deze eenheden-ID gaan van 0000_{hex} tot en met $FFFE_{hex}$. De waarde $FFFF_{hex}$ definieert een niet aanwezige eenheid. Dank zij deze code kan de startroutine van de PCI-BIOS alle PCI-eenheden in het systeem terug vinden. De header is in twee delen gesplitst: de eerste 16 bytes (00_{hex} tot en met $0F_{hex}$) zijn voor alle PCI-eenheden op dezelfde manier ingedeeld, de volgende 48 bytes kunnen voor de aanwezige PCI-eenheden een verschillende layout bevatten.
- Header:
  De samenstelling van de eenheden-ID wordt door het veld Header (offset $0E_{hex}$) bepaald. Op dit moment is alleen een enkel header-type gedefinieerd (Header = 0). Header-bit 7 geeft aan of het een multifunctionele eenheid is (bit 7 = 1) of een normale eenheid (bit 7 = 0).
- Maker-ID:
  De Maker-ID wordt door de commissie PCI-SIG, die de PCI-standaard heeft opgesteld, uitgedeeld.
- Klassecode:
  De Klassecode geeft aan om wat voor soort PCI-eenheid het gaat. Het veld is in drie eenheden van één byte ingedeeld. Het meest significante byte (offset $0B_h$) bepaalt de basisklassecode, het middelste byte (offset $0A_{hex}$) de subklassecode en het minst significante byte (offset 09_{hex}) bevat een programmeer-interface voor de betreffende eenheid.
- Opdracht:
  Het veld Opdracht in de header geeft informatie over de manier waarop een PCI-eenheid op PCI-cycli reageert. Wanneer men de waarde 0000_{hex} in het opdrachtregister zet, wordt de bijbehorende PCI-eenheid gedeactiveerd: deze reageert dan alleen nog maar up configuratiecycli.
- Status:
  Naast het opdrachtregister is er ook een statusregister, dat de status van de betreffende eenheid tijdens een PCI-handeling bevat.
- Cache-Line Grootte:
  Het header-veld CLG (Cache Line-grootte) definieert de grootte van de cache-line van het systeem in eenheden van 32 bytes. De waarde is bijvoorbeeld voor de opdracht geheugenschrijfbenadering met invalidatie nodig.
- Latentie:
  Het veld Latentie geeft aan, hoe lang een PCI-buscyclus kan duren: de effectieve tijdsduur bedraagt Latentie + 8 PCI-klokcycli.
- BIST:
  Dit is de Built In Self Test. Het meest significante bit in het BIST-register geeft aan of de eenheid een zelftest (built-in self-test) kan uitvoeren (Bit 7 = 1). Is dit het geval dan wordt de BIST oproepen door bit 6 de waarde 1 te geven. De BIST retourneert in de vier minst significante BIST-bits 0 tot en met 3 een exit-code. Een van nul verschillende waarde duidt op een fout.
- INT
  De INT bepaalt in een PC-AT naar welke hardware-interrupt lijn IRQx de interrupt-aansluiting van de eenheid of functie wordt vertaald. Geldige waarden zijn natuurlijk 0 tot en met 15, overeenkomend met IRQO tot en met IRQ15. De interrupt-routing logica van

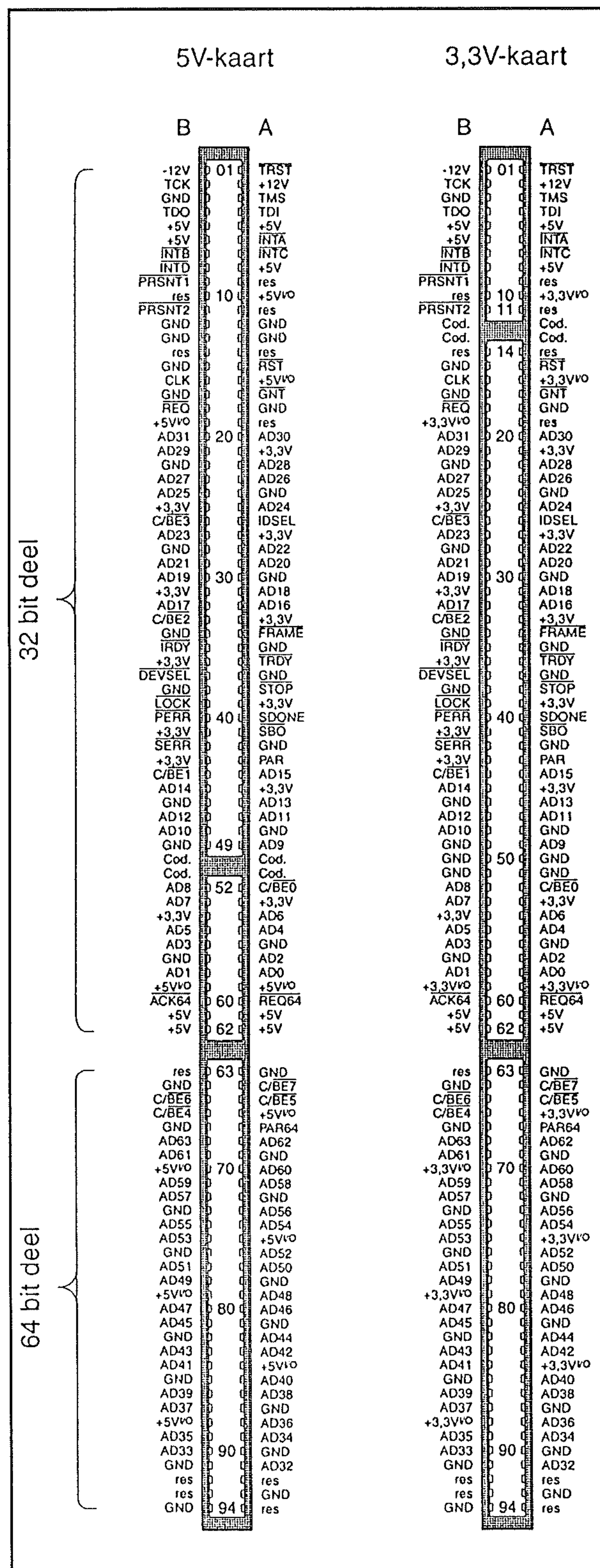
19.1 Architectuur van de computer hardware en de bus

de PCI-bridge activeert dan de juiste ingang van de daarvoor verantwoordelijke elektronica.
- MinGNT en MaxLat
  Deze beide read-only registers MinGNT en MaxLat geven de door de fabrikant van de PCI-eenheid gewenste minimale en maximale latentiewaarde aan, zodat de eenheid de PCI-bus optimaal kan benutten.

De basis- en subklassecodes
De basis- en subklassecodes zijn als volgt toegewezen:
- basiscode 00_{hex}:
  een dergelijke eenheid bestond al voordat de definiëring tot stand kwam.
  - subcode 00_{hex}:
    alles behalve VGA
- basiscode 01_{hex}:
  controller voor een massageheugen
  - subcode 00_{hex}:
    SCSI-controller
  - subcode 01_{hex}
    IDE-controller
  - subcode 02_{hex}:
    floppy-controller
  - subcode 03_{hex}:
    IPI-controller
  - subcode 80_{hex}:
    een andere controller
- basiscode 02_{hex}:
  een netwerk-controller
  - subcode 00_{hex}:
    Ethernet controller
  - subcode 01_{hex}:
    Token Ring controller
  - subcode 02_{hex}:
    FDDI controller
  - subcode 80_{hex}:
    een andere controller
- basiscode 03_{hex}:
  een video-controller
  - subcode 00_{hex}:
    VGA controller
  - subcode 01_{hex}:
    XGA controller
  - subcode 80_{hex}:
    een andere controller
- basiscode 04_{hex}:
  een multimedia-eenheid
  - subcode 00_{hex}:
    video kaart
  - subcode 01_{hex}:
    audio kaart
  - subcode 80_{hex}:
    een andere eenheid
- basiscode 05_{hex}:
  een geheugen-controller
  - subcode 00_{hex}:
    RAM controller
  - subcode 01_{hex}:
    flash-geheugen controller
  - subcode 80_{hex}:
    een andere controller
- basiscode 06_{hex}:
  een brug
  - subcode 00_{hex}:
    de host-brug
  - subcode 01_{hex}:
    ISA brug
  - subcode 02_{hex}:
    EISA brug
  - subcode 03_{hex}:
    MCA brug
  - subcode 04_{hex}:
    PCI-PCI brug
  - subcode 05_{hex}:
    PCMCIA brug
  - subcode 80_{hex}:
    een andere brug
- basiscode 07_{hex} tot en met FE_{hex}:
  gereserveerd

Het PCI-slot
De slot's van de PCI-structuur zijn niet te vergelijken met deze van de oudere bussystemen.

19.1 Architectuur van de computer hardware en de bus



Figuur 3/19.1-29: De slot's van de PCI-architectuur.

Het slot is, zonder compromissen te doen, samengesteld rond 32 bit brede adres- en data-bussen. Daarnaast is een uitbreiding voorzien naar 64 bit brede bussen. De samenstelling van het PCI-slot is getekend in figuur 3/19.1-29.

Zoals uit deze figuur blijkt, heeft het PCI-slot twee uitvoeringen: een voor uitbreidingskaarten die gevoed worden met +5 V en een voor de nieuwe energiezuinige kaarten (groene kaarten) die een voedingsspanning van slechts +3,3 V hebben. Door het aanbrengen van een nokje, respectievelijk tussen de contacten 49/52 (+5 V) en 11/14 (+3,3 V) kan een +3,3 V kaart nooit in een +5 V slot gestoken worden.

Energiebesparing

Een andere energiebesparende eigenschap van het PCI-slot is dat iedere PCI-eenheid geklokt moet kunnen worden met frequenties tussen 0 Hz en 33 MHz. Hierdoor kan de klok worden uitgeschakeld, hetgeen het opgenomen vermogen met ongeveer 99 % reduceert.

PRESENT

De twee PRESENT-contacten PRSNT1 en PRSNT2 worden door een PCI-kaart met 0 V of +5 V verbonden, om naast het aanwezig zijn van de kaart ook de vermogensopname van de elektronica op de kaart aan te kunnen geven. Is er geen kaart aanwezig, dan blijven de twee lijnen uiteraard open en liggen op de positieve voedingsspanning ("H"). De codering van deze contacten is:

- PRSNT1 = "H", PRSNT2 = "H":
  geen kaart aanwezig
- PRSNT1 = "L", PRSNT2 = "H":
  kaart met maximaal 25 W vermogensopname
- PRSNT1 = "H", PRSNT2 = "L":

kaart met maximaal 15 W vermogensopname

- $\overline{PRSNT1}$ = "L", $\overline{PRSNT2}$ = "L": kaart met maximaal 7,5 W vermogensopname

Universele kaarten

De met $XXX^{I/O}$ gecodeerde aansluitingen stellen de speciale voedingsaansluitingen van universele kaarten voor, die zowel met +3,3 V als met +5 V voeding kunnen werken. Dergelijke kaarten kunnen dus in een willekeurig slot worden geplaatst. Naast deze speciale voedingsaansluitingen zijn echter ook normale aansluitingen voor +5 V (5 V kaart) en +3,3 V (3,3 V kaart) aanwezig.

Het 64 bit deel van het slot

Het 64 bit deel is door een schot van het 32 bit deel gescheiden. Dit deel is niet verplicht en bevat het meest significante dubbelwoord van een vierdubbelwoord. De stuur- en statusaansluitingen $\overline{REQ64}$ en $\overline{ACK64}$, die de besturing van de 64 bit uitbreiding verzorgen, zitten wél in het 32 bit deel van het slot.

Compacte bouwwijze

Door de zeer dicht op elkaar liggende contacten is het volledige 64 bit PCI-slot zeer compact. Dankzij de adres/datamultiplexing heeft het 64 bit PCI-slot zelfs minder contacten dan een MCA-slot: 188 versus 202 aansluitingen.

Shared slot

Bij een zogenoemd "shared slot" is achter het normale PCI-slot nog een ISA-, EISA- of MCA-slot geplaatst. In principe kan een kaart namelijk zowel een PCI- als een ISA/EISA/MCA-interface bevatten en daardoor toepasbaar zijn in beide busstructuren. Er zijn echter weinig fabrikanten van moederborden en kaarten te vinden, die van deze optie gebruik maken. De normale gang van zaken is dat in het ISA/EISA/MCA-deel van een shared slot een echte ISA/EISA/MCA-kaart wordt geplaatst en in het PCI-deel een echte PCI-kaart.
Bovendien is een dergelijke structuur in feite overbodig. Ieder PCI-moederbord beschikt immers over een aantal "normale" EISA-sloten!

Boundary Scan Testing

Het PCI-slot ondersteunt de door de JTAG gestandaardiseerde testprocedure BST, oftewel Boundary Scan Testing. Hiervoor zijn twee aansluitingen aanwezig, TDO en TDI. PCI-kaarten die geen JTAG Boundary Scan Testing ondersteunen, moeten de aansluitingen TDO en TDI intern kortsluiten. Meestal wordt namelijk één enkele testlus over het gehele moederbord en alle PCI-slot's aangebracht. Dit maakt de testprocedure aanmerkelijk eenvoudiger, omdat er slechts één gemeenschappelijke ingang en slechts één gemeenschappelijke uitgang aanwezig is. Zouden dus op een PCI-kaart zonder test-elektronica de aansluitingen TDO en TDI niet zijn doorverbonden, dan zou de testlus worden onderbroken.

3/19.4

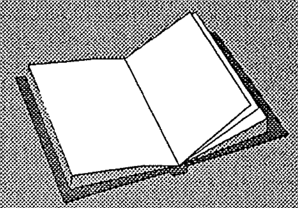
Principes van gegevensopslag op diskette

Inleiding

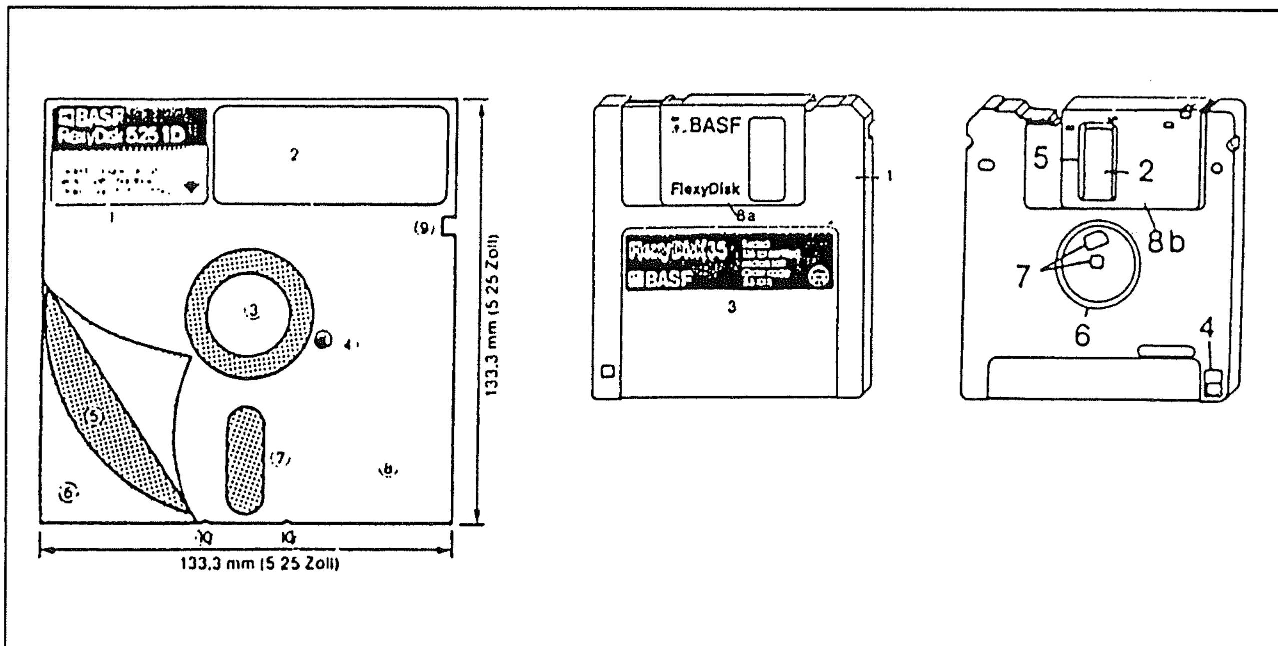
De magnetische zachte schijf
Het opslagmedium "magnetische zachte schijf", kortweg diskette genoemd, werd door IBM ontwikkeld en in eerste instantie onder de vorm van 8 inch (20,3 cm) schijven toegepast in grote computersystemen. Dit opslagmedium bleek een groot succes en dank zij uitvoerige research en verbetering van materialen en constructies werd het mogelijk de diskettediameter te verlagen tot 5.25 inch (13,3 cm) en later tot 3.5 inch (8,9 cm). Bij de introductie van de Personal Computer door IBM werd weliswaar een cassette-interface ingebouwd, maar was ook de mogelijkheid voorzien gebruik te maken van 5.25 inch schijven. Omdat niemand gebruik maakte van de cassette-interface werd deze uitbreiding al snel uit het ontwerp verwijderd, zodat als enig opslagmedium de zachte schijf ter beschikking stond. In eerste instantie werd gebruik gemaakt van enkelzijdige schijven met een opslagcapaciteit van ongeveer 160 kB. De allereerste versies van het besturingssysteem MS-DOS ondersteunden immers alleen dit formaat. Nadien werden de dubbelzijdige schijven ingevoerd en werd de informatiedichtheid opgevoerd, zodat op dit moment alle PC's werken met dubbelzijdige schijven met een capaciteit van 360 kB tot 1,44 MB. Diskettes met een capaciteit van 2,88 MB zijn beschikbaar, maar breken niet echt door.

Twee formaten
Lange tijd was de 5.25 inch diskette dé standaard van het PC-gebeuren. De kleinere 3.5 inch schijven werden echter door concurrerende systeemfabrikanten omarmd, zodat zowel IBM als alle cloonproducenten genoodzaakt waren over te schakelen naar dit systeem. Op dit moment is de 3.5 inch diskette dé standaard. In figuur 3/19.4-1 worden de twee disketteformaten voorgesteld.
Officieel wordt de 5.25 inch diskette een mini-diskette genoemd. Een logisch gevolg van het feit dat de allereerste schijven 8 inch groot waren! De eigenlijke schijf <5> is ondergebracht in een flexibele kunststofmantel <8>, de "jacket" die aan de binnenkant voorzien is van een beschermende en smerende laag <6>.

LEES OOK:
Hoofdstuk 3/19.5
Hoofdstuk 3/19.6



19.4 Principes van gegevensopslag op diskette



Figuur 3/19.4-1: De uiterlijke verschijning van een 5.25 en een 3.5 inch diskette.

Men noemt dit "het vlies" van de diskette. In de mantel is aan weerszijden een langwerpige opening <7> aangebracht, die de kop-opening wordt genoemd en waarlangs de kop van het loopwerk toegang krijgt tot de gevoelige laag op de schijf. In het midden is het centrale gat <3> aanwezig, waarmee de schijf wordt gekoppeld aan de as van de motor in de diskdrive. Naast dit gat bevindt zich het indexgat <4>. De positie van dit gat wordt in de drive uitgelezen door een optisch systeem, waardoor de besturing een referentie heeft voor het bepalen van de momentele positie van de schijf ten opzichte van de kop. In de mantel is een kleine uitsparing <9> aangebracht, die kan worden afgedekt met een stickertje en waardoor het systeem weet dat er alleen van de diskette gelezen mag worden. Dit is het bekende "write protect" beveiligingssysteem tegen het per ongeluk wissen of overschrijven van belangrijke informatie. In de rechter bovenhoek zit een sticker <I> met de fabrikant- en diskettegegevens, daarnaast is een ruimte <2> gereserveerd voor gebruikersinformatie.

De mini-diskette heeft een heleboel zwakke punten:

- de mantel is flexibel en geeft in feite maar erg weinig bescherming tegen vervormingen van de schijf;
- de gevoelige laag op de schijf is zonder meer aan te raken via de kop-opening <7>;
- zelfs als de schijf in het opbergmapje wordt opgeslagen kan stof tamelijk gemakkelijk doordringen tot de schijf via de kop-opening;
- de manier waarop de schijf wordt aangedreven is zonder meer uit de tijd te noemen, het centrale gat <3> is een zeer zwak punt van de constructie en kan gemakkelijk beschadigd worden;
- de manier waarop de diskette "write protected" wordt is gebruikersonvriendelijk.

De constructie van een 3.5 inch diskette is heel wat ingewikkelder, maar vangt alle bezwaren en zwakke punten van de mini-

diskette elegant op. De behuizing <1> bestaat nu uit star kunststof dat de schijf volledig beschermd tegen iedere mechanische mishandeling. In feite zou men zelfs van een 3 mm dikke "cartridge" kunnen spreken! De eigenlijke schijf <2> is in ongebruikte toestand volledig afgeschermd van de buitenwereld door middel van de afsluiting <8>. Deze metalen klep veert steeds naar de gesloten stand terug en vandaar wordt deze wel eens de "shutter" van de schijf genoemd. In de geopende stand wordt de schijf bereikbaar voor de kop via langwerpige openingen in de shutter en de cartridge, die in deze stand samenvallen. Het aandrijfsysteem bestaat uit een stevig op de diskette bevestigd metalen schijfje <6> dat bereikbaar is via een ronde opening in de cartridge. Op deze schijf zitten twee openingen <7> die gebruikt worden voor het centreren van de schijf op de as van de motor en voor het aandrijven van de schijf. Het "write protect" systeem bestaat uit een klein gaatje <4> in de hoek van de cartridge, dat door middel van een kunststof klepje gesloten kan worden. Als dit klepje open is kan de schijf alleen gelezen worden. De eigenlijke schijf bestaat uit een 80 µm dikke drager van polyester waarop men aan beide zijden een 2,5 µm dikke laag van een magnetisch gevoelige stof heeft aangebracht. In feite hetzelfde principe dat gebruikt wordt bij audio- of video-banden.

De gevoelige laag
De gevoelige laag van beide diskette-standaarden bestaat uit zeer fijn ijzeroxide-poeder, dat met oplos- en bindmiddelen wordt omgezet in een emulsie. Deze emulsie wordt op de polyesterfolie aangebracht waarna men het oplosmiddel laat verdampen en een mooie egale laag achterblijft. De ijzeroxyde deeltjes hebben een bepaalde magnetische polariteit. Na het opbrengen van de laag zijn deze polariteiten volgens een volledig willekeurig patroon gericht. Er is dus geen resulterend veld aanwezig, de schijf is magnetisch neutraal. De deeltjes kunnen echter magnetisch gepolariseerd worden. Dat wil zeggen dat onder invloed van een magnetisch veld alle magnetische deeltjes zich in dezelfde richting gaan opstellen als dat veld en nadien in deze stand blijven staan. Als men een magnetisch veld aanbrengt op een klein rechthoekje van de laag, dan zullen de deeltjes zich daar gaan richten met als gevolg dat er een resulterend magnetisch veld ontstaat. Dit veld kan nadien weer vrij eenvoudig gedetecteerd worden. Als dit gepolariseerd rechthoekje onder de kop draait zal het magnetisch veld dat rond het gepolariseerd oppervlak aanwezig is in de spoel van de kop een zeer kleine elektrische spanning induceren.

Betrouwbaarheid van diskettes

Inleiding
Af en toe verschijnen er in de pers, naast berichten over het hernieuwd opduiken van het monster van Loch Ness, verontrustende berichten waaruit zou blijken dat magnetische opslagmedia, zoals geluid- en videotapes maar ook computerdiskettes, zeer onbetrouwbaar zijn. Met name wordt in deze artikelen de betrouwbaarheid op lange termijn ter diskussie gesteld. Het is dus noodzakelijk dieper in te gaan op dit probleem. De betrouwbaarheid van een computerdiskette wordt bepaald door drie factoren:

19.4 Principes van gegevensopslag op diskette

- de mechanische levensduur;
- de magnetische stabiliteit;
- de lange-termijn archiveringszekerheid.

In de volgende kopjes worden deze drie factoren afzonderlijk besproken.

De mechanische levensduur

Deze grootheid bepaalt de termijn waarbinnen de op de diskette inwerkende mechanische krachten de goede werking nog niet nadelig beïnvloeden. Op de eerste plaats werken er mechanische krachten in op de positionerings-opening in het midden van de diskette. Deze krachten kunnen op lange termijn tot beschadigingen en vervormingen van dit gat leiden, waardoor de diskette niet goed gecentreerd wordt in het loopwerk en de kop(pen) de sporen op de schijf niet meer kunnen lezen. Op de tweede plaats worden er op het oppervlak van de schijf wrijvingskrachten uitgeoefend die groter worden naarmate de diskette vaker in en uit de drive wordt genomen. Zeker als de diskette ooit door ondeskundige behandeling werd geknikt kunnen deze krachten zo groot worden dan de aandrijving van de drive de randen van het positioneringsgat kan beschadigen. Een derde bron van mechanische beschadigingen is het ondeskundig inbrengen van de diskette in het loopwerk. Men moet de diskette steeds volledig horizontaal houden en langzaam met een vloeiende beweging in de gleuf duwen. Houdt men de diskette niet echt horizontaal, dan zal de magnetische schijf gebogen worden bij het inzetten, waardoor nadien grote wrijvingskrachten kunnen ontstaan. Ook bij de schijnbaar robuuste 3.5 inch diskettes kunnen deze problemen optreden!
De koppen worden bij het lezen of schrijven van informatie in contact gebracht met de zeer dunne magnetische laag. Te verwachten is dat er slijtage in deze laag optreedt en dat een diskette die vele malen is "afgespeeld" vormen van magnetisch verval gaat vertonen. Praktijkproeven die door vrijwel alle grote fabrikanten zijn uitgevoerd bewijzen dat dit erg meevalt. Rekening houdende met de omwentelingssnelheid van 300 toeren per minuut zou men een en hetzelfde spoor meerdere maanden continu kunnen laten uitlezen zonder dat er merkbare magnetische slijtage optreedt. De Duitse fabrikant BASF garandeert zijn diskettes voor meer dan 70.000.000 actieve omwentelingen! Dat komt overeen met 4.000 uur per spoor, een waarde die zelfs de meest actieve diskette ook na 20 jaar dagelijks gebruik nooit bereikt. Uiteraard gelden bovenstaande gegevens alleen als de diskettes worden gebruikt in volledig storingsvrij werkende loopwerken die goed onderhouden worden.

De magnetische stabiliteit

Het registratieproces komt, zuiver fysisch bekeken, neer op het magnetiseren van een deel van de gevoelige laag op de schijf. Alle microscopische magnetische deeltjes in een uiterst klein gebied van 4 bij 1,5 µm worden door het magnetische veld dat de kop opwekt in dezelfde richting gepolariseerd. Deze magnetische polarisatie is remanent hetgeen wil zeggen dat deze toestand blijft bestaan, ook nadat het veroorzakende magnetisch veld van de kop verwijderd is. Men kan zich de vraag stellen hoe stabiel dit uiterst kleine magneetje op de lange termijn is. Zoals bekend worden remanente magnetische velden afgebroken onder invloed van het weliswaar zeer zwakke maar duidelijk overal aanwezige magnetisch veld van de aarde. Daarnaast trillen moleculen en atomen onder in-

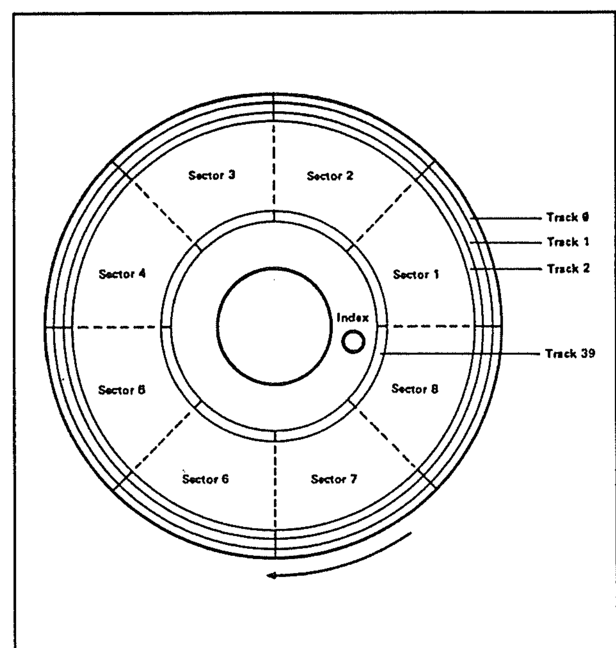
vloed van thermische energie. Door deze trillingen zullen de in één richting gepolariseerde magnetische deeltjes langzaam maar zeker weer een willekeurig richtingspatroon innemen. Volgens de technici van de diskettefabrikanten is het remanent magnetisch veld van een beschreven diskette zeer stabiel. Laboratoriumproeven waarbij beschreven diskettes kunstmatig verouderd werden en waarbij het meetbare verval van de gemagnetiseerde delen werd geëxtrapoleerd wijzen op een verwachte levensduur van 100 jaar. Vergeet daarbij niet dat het principe van de magnetische registratie zeer oud en dus zeer betrouwbaar is. De magneetband voor het registreren van geluid en spraak werd in 1936 ontwikkeld en de muziek die op deze allereerste banden werd opgenomen is nu nog steeds met dezelfde kwaliteit te beluisteren. Frequentietests in diverse laboratoria hebben dit bewezen!

De lange-termijn archiveringszekerheid
Bij het opslaan van diskettes in archieven moet men enige eenvoudige richtlijnen in acht nemen, die in feite ook al gelden voor het bewaren van papieren archieven. Diskettes moeten bij normale omgevingstemperaturen worden bewaard, dus tussen de +5 en de +25 °C. Het vochtigheidsgehalte van de lucht moet rond de 50 % liggen. Opslag in te droge ruimten heeft tot gevolg dat het smeermiddel dat in de diskettes is verwerkt uitdroogt en de mechanische wrijving van de in hun hoes ronddraaiende schijven toeneemt.

Formatteren van diskettes

Inleiding

Bij de aankoop van een diskette is het magnetisch oppervlak maagdelijk. Een zee van magnetische opslagruimte, die echter wél op een speciale manier voorbereid moet worden om er gegevens in op te kunnen bergen. Deze voorbereiding noemt men het formatteren van de diskette. Het formatteringsproces wordt uitgevoerd door het besturingssysteem van de computer in samenwerking met de disk-controller, een intelligent IC dat gebruikt wordt voor het besturen van de diskdrive. Door het formatteren wordt er als het ware een magnetische structuur op het oppervlak van de schijf aangebracht, waardoor de disk-controller en het besturingssysteem gemakkelijk toegang krijgen tot iedere willekeurige plek van de schijf. Op deze manier kunnen gegevens op een zeer snelle manier op de schijf opgezocht worden en kan het systeem de schijf op een economische manier beheren. Door het formatteren krijgt de schijf een bepaalde structuur, een formaat dat wordt opgelegd door het besturingssysteem.



Figuur 3/19.4-2: De standaard indeling van een diskette in tracks en sectoren.

19.4 Principes van gegevensopslag op diskette

Basis-indeling

Ondanks het feit dat er verschillende formaten in omloop zijn hebben deze allemaal een identieke basis-struktuur. Zoals blijkt uit figuur 3/19.4-2 wordt de schijf op de eerste plaats ingedeeld in een aantal sporen of tracks. Deze tracks vormen concentrische cirkels op het oppervlak van de schijf met steeds kleiner wordende stralen. Het aantal tracks is afhankelijk van het formaat dat men (of het besturingssysteem) gekozen heeft. De tracks worden genummerd, waarbij het spoor dat het dichtst bij de rand van de schijf ligt nummer 0 krijgt toegewezen. Om de track-omtrek niet al te zeer te laten dalen met het kleiner worden van de track-straal wordt slechts een zône met een breedte van ongeveer 2,5 cm werkelijk gebruikt. Het zal duidelijk zijn dat deze nuttige band zo dicht mogelijk tegen de rand van de schijf ligt. Daar is immers de straal van de tracks het grootst en bijgevolg ook de omtrek van de tracks. Het grootste deel van het oppervlak van de schijf wordt dus niet gebruikt!

Ieder track wordt bovendien verdeeld in een aantal segmenten of sectoren. De fysische lengte van de sectoren wordt steeds kleiner naarmate het track-nummer stijgt, omdat de omtrek van de tracks kleiner wordt als de straal daalt. Ook de sectoren worden genummerd, waarbij echter het eerste segment, voor de broodnodige afwisseling, nu het nummer 1 toegevoegd krijgt.

Er bestaat een bepaald verband tussen de positie van het indexgat in de schijf en de plaats waar het eerste segment start. Let bovendien op de relatie tussen de nummervolgorde en de draairichting van de schijf. De sector is de kleinste te benaderen eenheid van de schijf. Alle gegevens worden sectorgewijs op de schijf geschreven, waarbij uiteraard in de meeste gevallen meerdere sectoren nodig zijn om één file te bevatten. Omdat iedere sector en iedere track een nummer heeft kan het besturingssysteem in samenwerking met de disk-controller de koppen van de disk-drive naar ieder gewenst segment van de schijf sturen. Het volstaat immers een bepaalde codering in ieder segment op te nemen waaruit duidelijk blijkt de hoeveelste sector van het hoeveelste track dit is. Ook de zijde van de diskette speelt hierbij natuurlijk een rol. Vandaar dat ook de zijden worden gecodeerd. De ene zijde krijgt de code 0, de andere code 1.

Slack

De manier waarop een diskette is ingedeeld in track's en sectoren heeft een belangrijke consequentie. Een bestand met een omvang van 1 byte zal toch op een diskette een volledige sector in beslag nemen!

De rest van de sector blijft dan leeg en is niet meer afzonderlijke benaderbaar vanuit het besturingssysteem. Zelfs als het besturingssysteem meldt dat een diskette helemaal vol is, zal er nog een heleboel plaats open zijn, die zit in sectoren die niet helemaal met een bestand gevuld zijn. Dit verschijnsel noemt men de "slack" van de diskette. Er zijn programma's die in staat zijn alle bestanden op een diskette tot één groot bestand samen te voegen en nadien dit ene bestand in aaneengesloten sectoren te schijven. Het zal duidelijk zijn dat een diskette hierdoor veel efficiënter gevuld wordt en er veel meer gegevens op passen.

De MS-DOS formaten

Wie "PC" zegt, denk "MS-DOS". Het aloude besturingssysteem van MicroSoft, vaak vervloekt maar nog steeds onmisbaar. De

19.4 Principes van gegevensopslag op diskette

hoeveelheid informatie die op een diskette kan worden weggeschreven wordt in eerste instantie nog steeds bepaald door dat besturingssysteem. De fundamentele indeling van een schijf in een aantal tracks die ieder een gelijk aantal sectoren bevatten treft men bij ieder formaat aan. Wat verschilt is het aantal tracks en het aantal sectoren per track. In figuur 3/19.4-3 is een historisch overzicht gegeven van de acht formaten die door de verschillende versies van MS-DOS ondersteund werden en worden.

CODERING	ZIJDEN	SECTOREN	TRACKS	CAPACITEIT
S-8	1	8	40	160 kB
D-8	2	8	40	320 kB
S-9	1	9	40	180 kB
D-9	2	9	40	360 kB
DD-9	2	9	80	720 kB
DD-15	2	15	80	1,2 MB
DD-18	2	18	80	1,44 MB
DD-36	2	36	80	2,88 MB

Figuur 3/19.4-3: De acht diskette-formaten die door MS-DOS ondersteund worden.

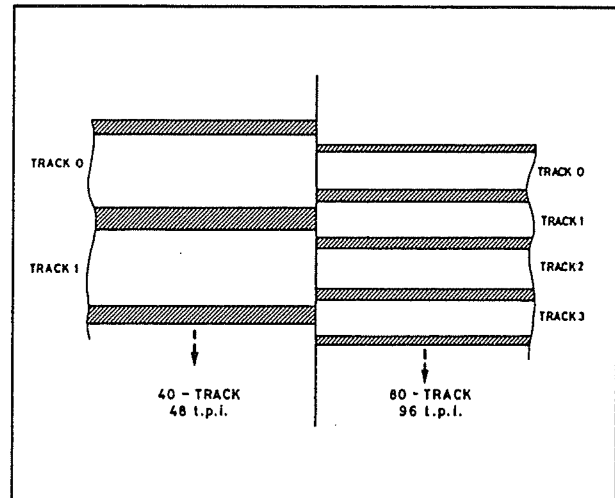
Een korte toelichting.

- S-8:
  Deze schijfstructuur was de enige die door versie 1.00 van MS-DOS werd ondersteund. Dit systeem werkt met enkelzijdige schijven, waarop 40 tracks worden geschreven die ieder uit 8 sectoren bestaan.
- D-8:
  De dubbelzijdige schijf werd ingevoerd bij versie 1.10. Uiteraard verdubbelt de capaciteit van de schijf hierdoor van 160 naar 320 kB.
- S-9 en D-9:
  Vanaf versie 2.00 werd een track ingedeeld in negen sectoren en omdat iedere sector een standaard hoeveelheid informatie kan bevatten nam de capaciteit van een schijf hierdoor met 12,5 %. toe.
- DD-9:
  Bij deze schijfstructuur worden er 80 tracks met 9 sectoren op het schijfoppervlak geschreven. Omdat deze dubbele hoeveelheid sporen op hetzelfde oppervlak moet worden ondergebracht moet de breedte van een track drastisch verlaagd worden.
- DD-15:
  Dit formaat wordt ondersteund door alle MS-DOS versies vanaf 3.0 en het is het bekende formaat waarmee high density 5.25 inch diskettes worden geformatteerd. Het aantal sectoren per track wordt verhoogd tot 15, waardoor de capaciteit van de schijf stijgt tot niet minder dan 1,2 MB. Er worden vanwege de zeer smalle tracks en de hoge informatiedichtheid vanwege de 15 sectoren zeer hoge eisen gesteld aan de oppervlaktestructuur van de schijf en men moet dan ook gebruik maken van de zogenaamde "high density" schijven.
- DD-18:
  Dit is het standaard formaat voor 3.5 inch high density diskettes, op dit moment dé standaard op alle systemen.
- DD-36:
  Een relatief nieuw formaat, waarmee men 2,88 MB op een standaard high density 3.5 inch diskette kan onderbrengen. Ondanks het feit dat MS-DOS dit formaat volledig ondersteunt zijn er maar weinig systemen te koop waarmee men dergelijke diskettes kan formatteren. Een en ander heeft ongetwijfeld te maken met het feit dat de standaard disk-controllers deze structuur niet ondersteunen en met het feit dat de ver-

spreiding van CD-ROM drivers, CD-Recorders en tape-streamers een dergelijk formaat in feite overbodig maken.

De track-dichtheid

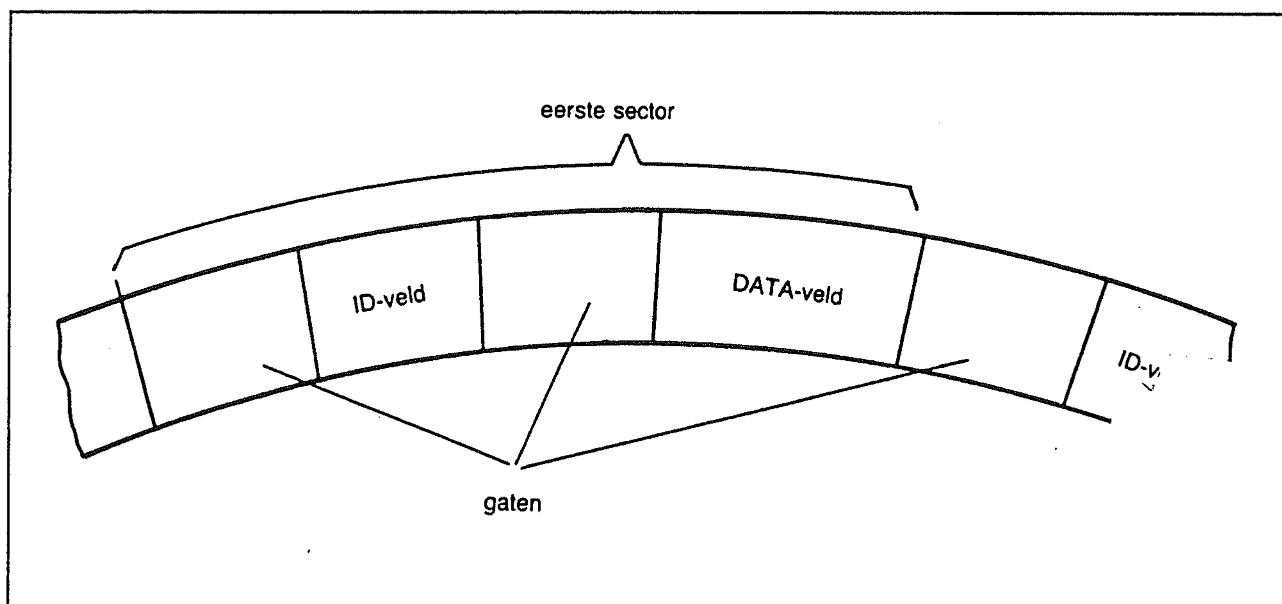
Het grote verschil tussen diskettes die geformatteerd zijn op een oeroude PC en deze die een soortgelijke behandeling op een modern systeem hebben ondergaan is dus het aantal tracks op de schijf. Het bruikbare oppervlak op de schijf is iets minder dan 1 inch (2,54 cm) breed, zodat de sporen erg dicht op elkaar liggen. Voor 40 track schijven is de track-dichtheid gelijk aan 48 tpi (tracks per inch), voor 80 track schijven loopt deze grootheid op tot 96 tpi. De relatie tussen de track-ligging op 40 en 80 track schijven is getekend in figuur 3/19.4-4. De tracks worden op de DD-schijven zo geplaatst dat iedere track met een even nummer precies in het midden ligt van de tracks op de schijf met lage dichtheid. De oneven tracks op de high density schijven overlappen twee opeenvolgende tracks van de low density disks.



Figuur 3/19.4-4: De relatie tussen de track-ligging op schijven met lage en hoge dichtheid.

Gegevensdichtheid

Computerinformatie bestaat uit bytes en het is dus logisch dat deze bytes ook als basis worden genomen voor het registreren van gegevens op een schijf. Bij de PC worden bij alle formaten 512 bytes per sector geschreven en gelezen.



Figuur 3/19.4-5: Het indelen van een sector in vier zônes tijdens het formatteren.

19.4 Principes van gegevensopslag op diskette

Hoe het mogelijk dat dit aantal zowel in een lange sector van een S-8 of een S-9 past als in de veel kortere sector van een DD-schijf wordt later toegelicht.

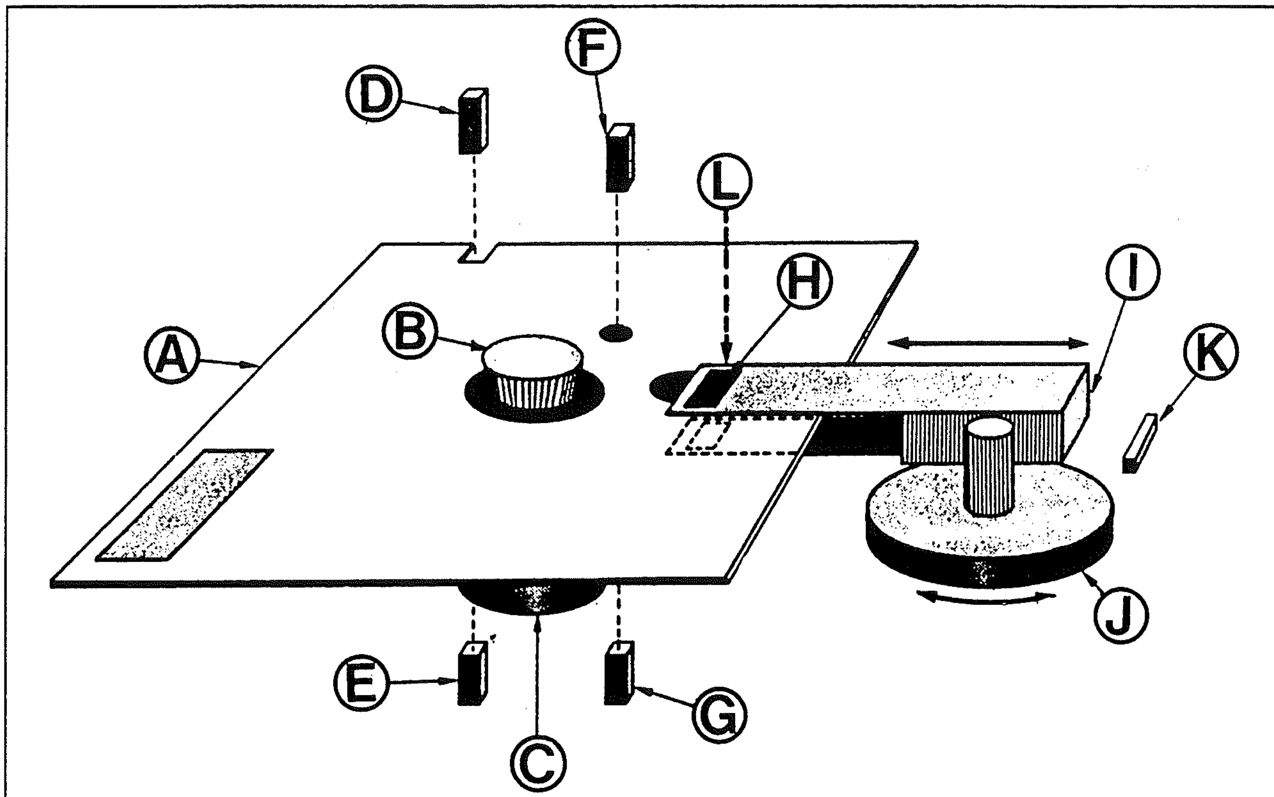
Herkenning van tracks en sectoren
De belangrijkste taak van het formatterings-proces is aan iedere sector op de schijf een ondubbelzinnige code toe te kennen, zodat het bedrijfssysteem deze sector snel kan terug vinden. Deze zogenaamde ID-code is opgebouwd uit de nummers van de schijfzijde, de track en de sector. De samenstelling van deze numerieke code is vrij ingewikkeld en doet verder weinig ter zake. In figuur 3/19.4-5 is getekend hoe de ruimte in een sector in vier zônes wordt ingedeeld. Een sector begint met een zogenaamd "gat", een ruimte die alleen wordt gebruikt voor het registreren van bepaalde synchronisatiepulsen (zie later). Nadien volgt het ID-veld, waarin onder meer de ID-code van de sector wordt genoteerd, maar ook andere gegevens die voor het systeem van belang zijn. Nadien volgt weer een "gat". De ruimte na dit tweede "gat" is vrij voor het opnemen van gegevens in het DATA-veld.

Het loopwerk, de interface en de disk-controller

Functies van het loopwerk
Het loopwerk, de officiële Nederlandse benaming voor het apparaat dat men meestal diskdrive of floppystation noemt, heeft een groot aantal taken te vervullen. Het is dan ook een ingewikkeld en precies apparaat en het is een wonder dat loopwerken tegenwoordig al voor minder dan 100 gulden te koop zijn en zo ongelooflijk betrouwbaar blijven werken! Een overzicht van de functies:
- het precies centrisch en zonder slip positioneren van de diskette rond de centrale as van de motor;
- het zo snel mogelijk versnellen van de schijf tot het juiste toerental;
- het zo constant mogelijk houden van dit toerental;
- het detecteren of de diskette "write protected" is;
- het detecteren van de positie van het indexgat;
- het detecteren van de positie van het eerste track op de schijf;
- de koppen snel naar het juiste track sturen;
- de koppen in contact brengen met de schijf;
- de gegevens in- of uitlezen zonder dat gegevens op naburige tracks worden beïnvloed.

In figuur 3/19.4-6 is de schematische opbouw van een loopwerk getekend. De diskette <A> wordt tussen de as van de motor <C> en een conische meenemer <B> geplaatst. Bij het sluiten van het loopwerk, door het omdraaien van het handeltje of het indrukken van een knop, wordt de conische meenemer in het centrale gat van de schijf geduwd. Dit proces is toegelicht aan de hand van figuur 3/19.4-7. De conische vorm van de meenemer zal er voor zorgen dat de schijf zuiver gecentreerd wordt rond de as van de motor, zodat de tracks gecentreerd op de schijf worden geschreven. Dit is zeer belangrijk omdat een schijf, dit in tegenstelling tot bijvoorbeeld een compact disk, geen enkel controlesysteem heeft dat er voor zorgt dat de koppen de tracks blijven volgen.

19.4 Principes van gegevensopslag op diskette

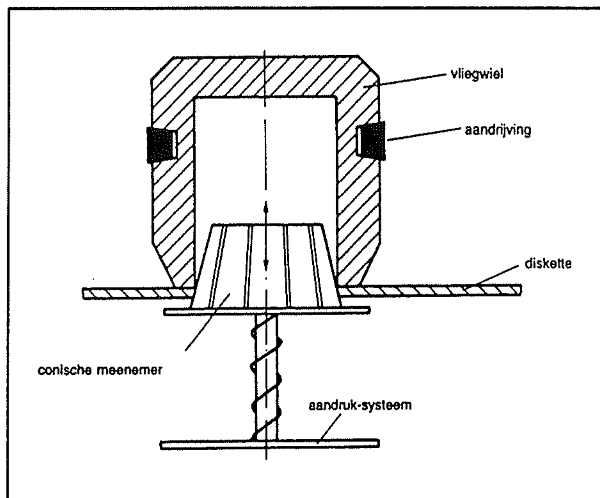


Figuur 3/19.4-6: De samenstelling van een loopwerk, in dit geval een voor 5.25 inch diskettes.

De minste of geringste afwijking van de centrering van de schijf rond de as van de motor heeft tot gevolg dat de schijf niet meer uitgelezen kan worden!
Terug nu naar figuur 3/19.4-6. Nadat de schijf is aangebracht zal een LED <D>, een smalle bundel licht werpen op de nok die afgedekt kan worden en die aangeeft of er op de schijf alleen gelezen of ook geschreven mag worden. Is de nok niet afgedekt, dan zal de lichtbundel de fotogevoelige detector <E> bereiken en daar een spanning in opwekken. Is dit signaal afwezig, dan weet de drive dat schrijfoperaties niet zijn toegestaan. De motor van het loopwerk is tegenwoordig direct gekoppeld met de as. Vroeger gebruikte men vaak snaaroverbrengingen zoals bij kassetterecorders, maar het zal duidelijk zijn dat deze systemen zeer slijtagegevoelig zijn. Bij het activeren van de drive zal de motor gaan draaien. Het is uiteraard belangrijk dat de motor de schijf zo snel mogelijk op het gestandaardiseerde toerental van 300 omwentelingen per minuut brengt. Alleen dan kan de schijf beschreven of uitgelezen worden. Vandaar dat men in de meeste gevallen een heel licht vliegwiel gebruikt en het stabiliseren van het toerental aan een elektronisch, kristalgestuurd systeem overlaat. Een typische stabiliseringstijd van het toerental is 250 ms. Boven het indexgat is een tweede LED <F> aangebracht, die een smalle lichtbundel op de juiste cirkeldiameter van de schijf werpt. Iedere keer als het indexgat onder deze LED draait zal de lichtbundel door het gat schijnen en de fotogevoelige detector <G> even belichten. Er wordt een puls opgewekt waardoor het systeem weet dat op dat moment het indexgat onder de LED doordraait. Op deze manier is het

19.4 Principes van gegevensopslag op diskette

systeem in staat te detecteren wanneer de eerste sector van een track precies onder de koppen zit.



Figuur 3/19.4-7: Het zuiver centreren van de schijf in de diskette rond de as van de motor door middel van een conische meenemer.

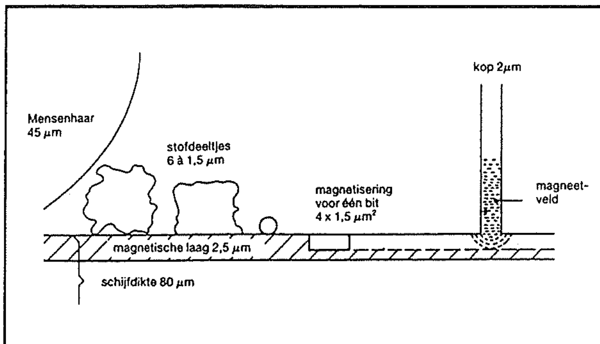
De kop of de koppen <H> zijn aangebracht aan weerszijden van een dunne vorkconstructie <I>. Deze vork kan radiaal heen en weer bewegen in de smalle opening die in de omhulling van de schijf is uitgespaard. Op deze manier kunnen de koppen in contact komen met het magnetische oppervlak van de schijf. Dat heen en weer bewegen van de kopconstructie wordt verzorgd door een stappenmotor <J>. Zo'n motor wordt gestuurd met pulsen. Bij een bepaalde pulscombinatie zal de as van de stappenmotor over een bepaald aantal graden gaan draaien. Door middel van een mechanische overbrenging wordt deze radiale beweging overgebracht op de vork, zodat deze bij iedere aansturing van de motor de koppen één track over de schijf verplaatst. Een van de belangrijkste gegevens van een diskdrive in dit kader is de zogenaamde staptijd, ook wel eens track-tot-track tijd genoemd. Deze tijd geeft aan hoe lang de constructie er voor nodig heeft om de koppen van het ene naar het volgende track te verplaatsen. Deze tijd ligt ergens tussen de 2 en de 30 ms. Zoals later zal blijken moet de kop bij het in- of uitlezen van een file vaak heen en weer springen over de schijf om vrije sectoren op te zoeken. Het zal dus duidelijk zijn dat deze staptijd de snelheid van de drive in hoge mate bepaalt!

Tot slot moet de constructie in staat zijn de koppen precies boven het eerste track te plaatsen. Bij het opstarten van de computer werkt de besturing immers nog niet volledig maar toch moet het systeem in staat zijn bepaalde gegevens, die op het eerste track aanwezig zijn, uit te lezen. Vandaar dat een zeer nauwkeurig werkende microschakelaar <K> is aangebracht, die wordt gesloten als de vork <I> zo ver naar buiten wordt verplaatst dat de koppen precies boven het eerste track staan.

Het koppensysteem

Wie zich de koppen van een diskdrive ongeveer voorstelt als deze van een kassetterecorder, alleen hoogstens wat kleiner, doet er goed aan figuur 3/19.4-8 eens aandachtig te bekijken! In deze figuur worden de afmetingen van enige typische schijfvervuilers op schaal vergeleken met de afmetingen van de kop. Het zal nadien wel onmiddellijk duidelijk zijn dat niet voor niets veel aandacht moet worden besteed aan het schoon houden van het oppervlak van de schijf! Toch zitten er in die 2 μm brede constructie niet minder dan drie koppen. De middelste is de eigenlijke schrijf/leeskop, die aan weerszijden wordt afgesloten door wiskopjes. Deze zorgen ervoor dat de ruimte tussen de tracks volledig vrij blijft van magnetische informatie.

19.4 Principes van gegevensopslag op diskette



Figuur 3/19-4-8: De onvoorstelbaar kleine afmetingen van de koppen krijgen reële betekenis als zij vergeleken worden met de afmetingen van enige schijfvervuilers.

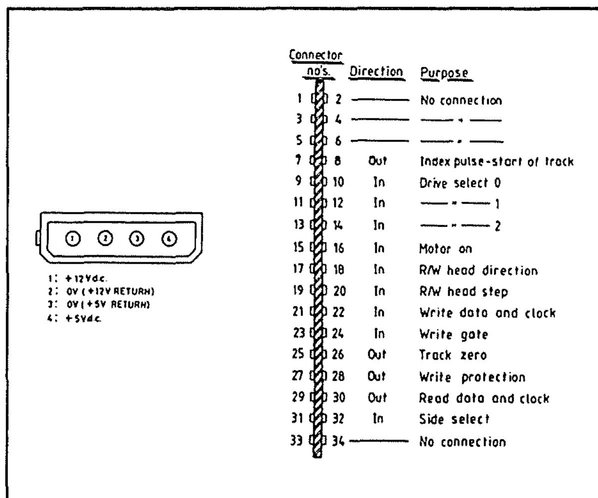
De koppen worden alleen in contact met de schijf gebracht als er echt gelezen of geschreven gaat worden. Bij de meeste systemen worden de twee tanden van de vork-achtige constructie waarop de koppen zijn gemonteerd door middel van een elektromagneet naar elkaar toe getrokken, waardoor de koppen tegen de schijf worden geperst.

De elektronica in de diskdrive
De schakelingen die ingebouwd zijn in de behuizing van de diskdrive zorgen voor het bedienen van de motoren en elektromagneten, het stabiliseren van het toerental van de motor, het opwekken van de stuurstroom voor de koppen, het versterken van het kopsignaal bij lezen, de schakelingen voor het registreren van de index en de "write protect" en het positioneren van de kop boven het eerste track. Meestal is ook nog een identificatieschakeling aanwezig, die aan de diskdrive een bepaald nummer toekent. Op deze manier kan men verschillende diskdrives parallel op dezelfde interface-kabel aansluiten. Zoals dadelijk zal blijken bevat de standaard interface tussen de drive-elektronica en de disk-controller drie "DRIVE SELECT" signalen, zodat alleen die drive geactiveerd wordt waarvan het identificatie-nummer overeen komt met het aangestuurde SELECT-signaal. Bovendien bevat de elektronica in de drive interface-schakelingen die alle uitgaande signalen op TTL-niveau brengen, zodat de volgens de TTL-technologie opgebouwde disk-controller deze signalen zonder meer kan verwerken.

De Shugart-interface
De schakelingen in het loopwerk worden tegenwoordig volgens een gestandaardiseerd systeem met de disk-controller, die meestal aanwezig is op een kaart of deel uitmaakt van het moederbord, verbonden. Deze interface, genoemd naar de fabrikant Shugart die deze connector het eerst toepaste, bestaat uit een randconnector op de dubbelzijdige print van de drive. Deze connector heeft 34 aansluitingen waarvan er echter slechts 17 worden gebruikt. Daarnaast is nog een eenvoudige vierpolige steker aanwezig die de voedingspanningen en de massa aan de diskdrive aanbiedt.
De codering van deze twee connectoren is getekend in figuur 3/19.4-9.
Een korte bespreking van de functie van de signalen.

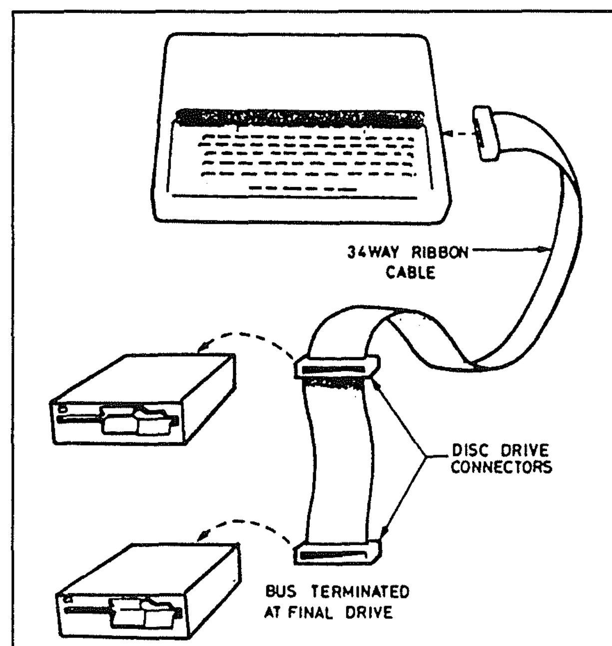
- DRIVE SELECT:
  Deze signalen zorgen er voor dat men verschillende loopwerken door middel van één kabel met aftakkingen op de disk-controller kan aansluiten. Dit systeem is op een zeer tot de verbeelding sprekende manier verduidelijkt in figuur 3/19.4-10. Het activeren van een SELECT-signaal heeft tot gevolg dat de koppen in de betreffende eenheid op de schijf worden gezet en deze drive klaar staat om te lezen of te schrijven.

19.4 Principes van gegevensopslag op diskette



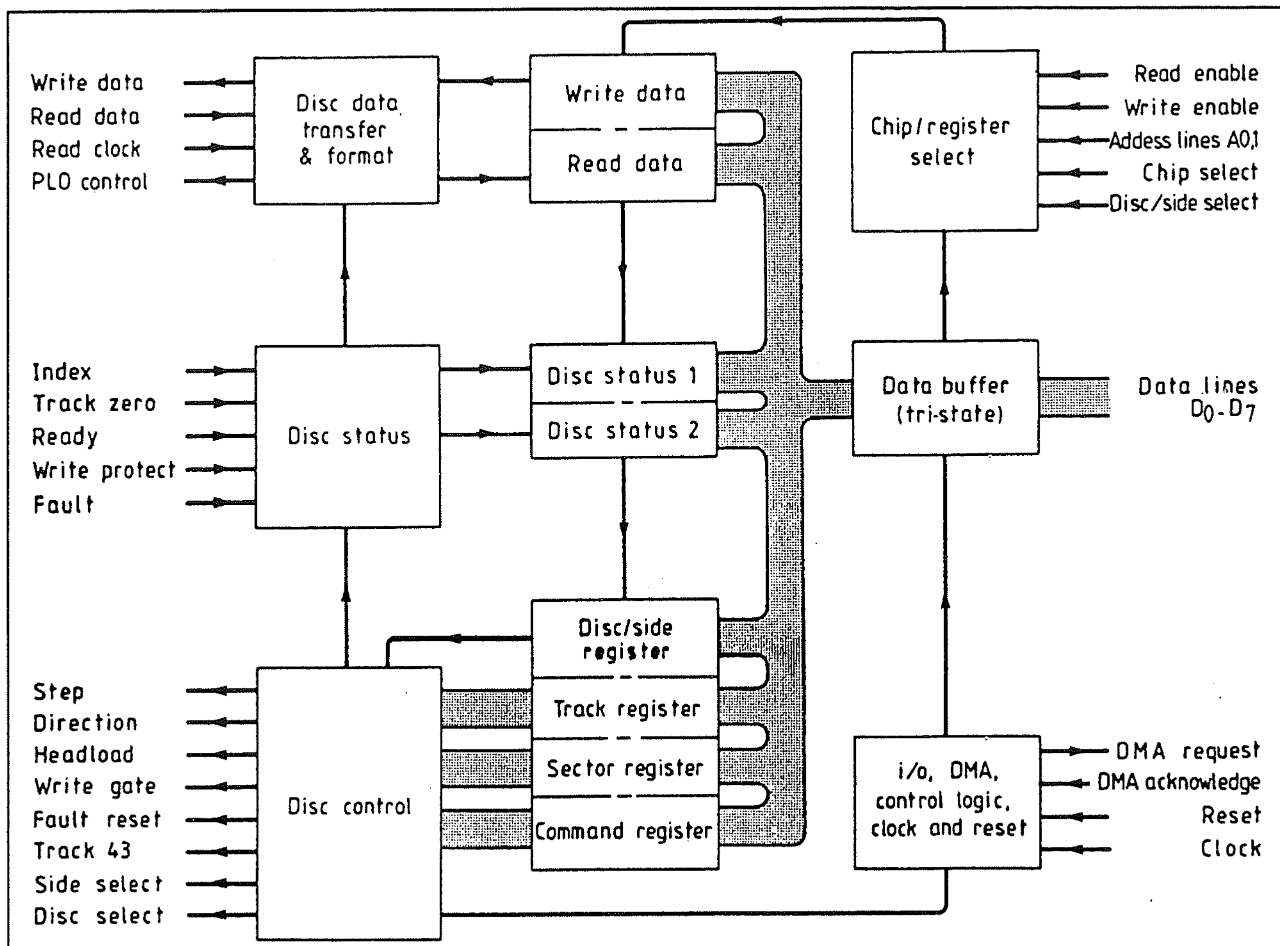
Figuur 3/19.4-9: De aansluitgegevens van de voeding-connector en van de edge-connector volgens de Shugart-norm.

- MOTOR ON:
  Stuurt de schijfmotoren in alle aangesloten drives aan, dus niet alleen in de geselecteerde drive! Dit is gedaan om een zeer snelle overschakeling van de ene naar de andere drive mogelijk te maken. Alle overige signalen werken alleen in op de geselecteerde drive, de niet geselecteerde stations zetten deze in- of uitgangen in een niet actieve tri-state toestand.
- R/W HEAD DIRECTION:
  Bepaalt de richting waarin de kopconstructie zal verplaatst worden.
- R/W HEAD STEP:
  Een puls op deze lijn verplaatst de kop één track in de door het vorige signaal bepaalde richting.
- WRITE DATA AND CLOCK:
  Verzendt de seriële pulsstroom die van de disk-controller afkomstig is en die naar de kop wordt gestuurd voor het schrijven van de informatie.
- WRITE GATE:
  Activeert de schrijf-elektronica in de drive.
- SIDE SELECT:
  Selecteert de diskzijde waarop geschreven of waaruit gelezen wordt.
- INDEX PULSE:
  Verschijnt uiteraard iedere keer als het indexgat onder de lichtgevende diode doordraait.
- TRACK ZERO:
  Geeft aan de disk-controller door dat de koppen precies boven het eerste track staan.
- WRITE PROTECTION:
  Wordt hoog als een niet te beschrijven schijf in de drive is geplaatst.
- READ DATA AND CLOCK:
  Stuurt de door de kop in lees-mode opgepikte en nadien versterkte signalen van de schijf naar de disk-controller.



Figuur 3/19.4-10: Het cascaderen van verschillende drives met behulp van een kabel.

19.4 Principes van gegevensopslag op diskette



Figuur 3/19.4-11: Het blokschema van een diskcontroller.

De disk-controller

Het intelligente deel van de elektronica wordt tegenwoordig meestal door één IC verzorgd, de zogenaamde disk-controller, geïntegreerd op het moederbord. Deze werkt als interface tussen de computerbus en de diskette-besturing en ontlast de centrale processor van de computer van een aantal taken. Het typische blokschema van een disk-controller is getekend in figuur 3/19.4-11. De rechter aansluitingen gaan naar de bus van de computer, de linker rechtstreeks of via tussenschakelingen naar de diskdrive. Het zal duidelijk zijn dat de disk-controller nauw samenwerkt met de centrale processor en het besturingssysteem.

Instructies en gegevens worden eerst in interne registers geplaatst en nadien overgedragen naar de databus van de computer.
Na het uitvoeren van een door de processor opgedragen instructie zal de controller een statusrapport geven, waaruit het besturingssysteem kan afleiden of alles naar wens is verlopen.
Het gaat in het kader van dit hoofdstuk te ver om diep in te gaan op de werking van deze schakeling. Vermeld wordt slechts dat de disk-controller in staat is een groot aantal handelingen zelfstandig te verrichten, waaronder:
- het resetten van de koppen op de eerste track;

19.4 Principes van gegevensopslag op diskette

- het uitschakelen van de motor als er binnen de vijf seconde geen schrijf- of leesinstructies gegeven worden;
- automatisch bepalen in welke richting de koppen verplaatst moeten worden om van een bepaalde track naar een andere te gaan;
- bij het opstarten van het systeem de koppen automatisch boven de eerste track zetten en de functie van de drive omschakelen naar lezen, zodat het systeem kan onderzoeken of de opstartfiles van het bedrijfssysteem op de diskette staan;
- het automatisch controleren van de geschreven gegevens op fouten en desgewenst het schrijfproces herhalen;
- het administreren van de positie van de kop;
- het aanvragen van een interrupt bij de processor.

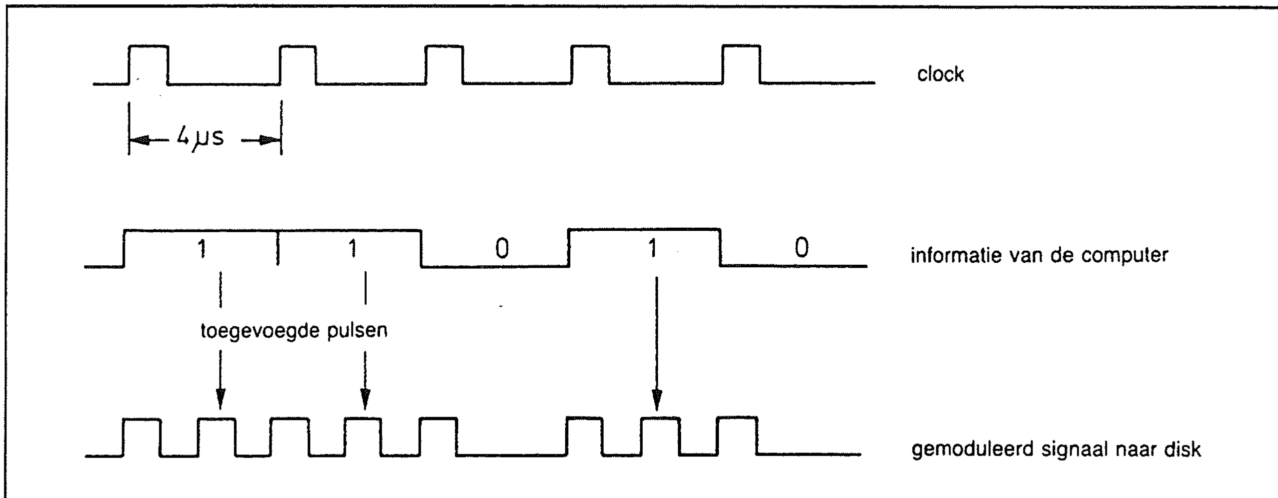
Daarnaast is de disk-controller in staat statussignalen te genereren uit de toestand van de schijf en deze aan de processor mede te delen. Dit zijn onder andere:
- schijf is klaar voor het ontvangen van de volgende instructie;
- schijf is "write protected";
- de door het bedrijfssysteem opgevraagde sector is niet aanwezig;
- geselecteerde sector kan niet beschreven worden;
- klaar voor het ontvangen van het volgend byte.

Het opslaan van gegevens

Magnetische registratie

De gegevens die op de schijf opgenomen moeten worden zijn in de computer beschikbaar onder de vorm van elektrische signalen, die ofwel 0 V ("L") of ongeveer +4,5 V ("H") kunnen zijn. Deze signalen kunnen niet als dusdanig op de schijf worden gezet, omdat er eerst natuurlijk een omzetting naar magnetische grootheden noodzakelijk is. Dat gaat vrij eenvoudig door voor ieder "H"-bit een korte stroomstoot door het zeer kleine lees/schrijfspoeltje in de kop te sturen. Deze stroom wekt een magnetisch veld op, dit veld magnetiseert een uiterst klein gebiedje van de dunne magnetische laag op het oppervlak van de schijf. De gegevens zitten echter in parallelle vorm in het geheugen van de computer. Een byte bestaat uit acht bits, die door hun momentele "L"/"H"-combinatie de inhoud van het byte vertegenwoordigen. Deze parallelle gegevens moeten dus eerst omgezet worden naar een seriële vorm, zodat de hoge pulsen uit het byte een na een opeenvolgende deeltjes van het track op de schijf kunnen magnetiseren. Bij dit proces is het echter noodzakelijk dat er een extra signaal wordt toegevoegd, het zogenaamde kloksignaal. Er worden namelijk geen spaties opgenomen tussen de opeenvolgende bits en uit opeenvolgende hoge bits zouden dus tot gevolg hebben dat er een breed gebied werd gemagnetiseerd in plaats van twee duidelijk van elkaar te onderscheiden magnetische zônes. Daarom wordt het begin van ieder bit, hoog of laag, geïdentificeerd door een smalle klokpuls die ook op de schijf als magnetische zône wordt geregistreerd. Zelfs bij het schrijven van een groot aantal opeenvolgende lage bits zal er toch een signaal op de schijf verschijnen en dit signaal kan bij het lezen van de informatie gebruikt worden om de seriële datastroom van de leeskop weer gesynchroniseerd om te zetten in parallelle informatie. Er bestaan twee systemen om de datagegevens te koppelen aan de klokpulsen.

19.4 Principes van gegevensopslag op diskette



Figuur 3/19.4-12: De magnetische registratie van gegevens volgens het FM-systeem.

Deze worden FM en MFM genoemd. Het FM-systeem wordt gebruikt bij alle schijven die geformatteerd zijn met 8 of 9 sectoren per track. Het plaatsbesparende MFM-systeem wordt toegepast bij alle DD-schijven, die meer dan 9 sectoren per track hebben.

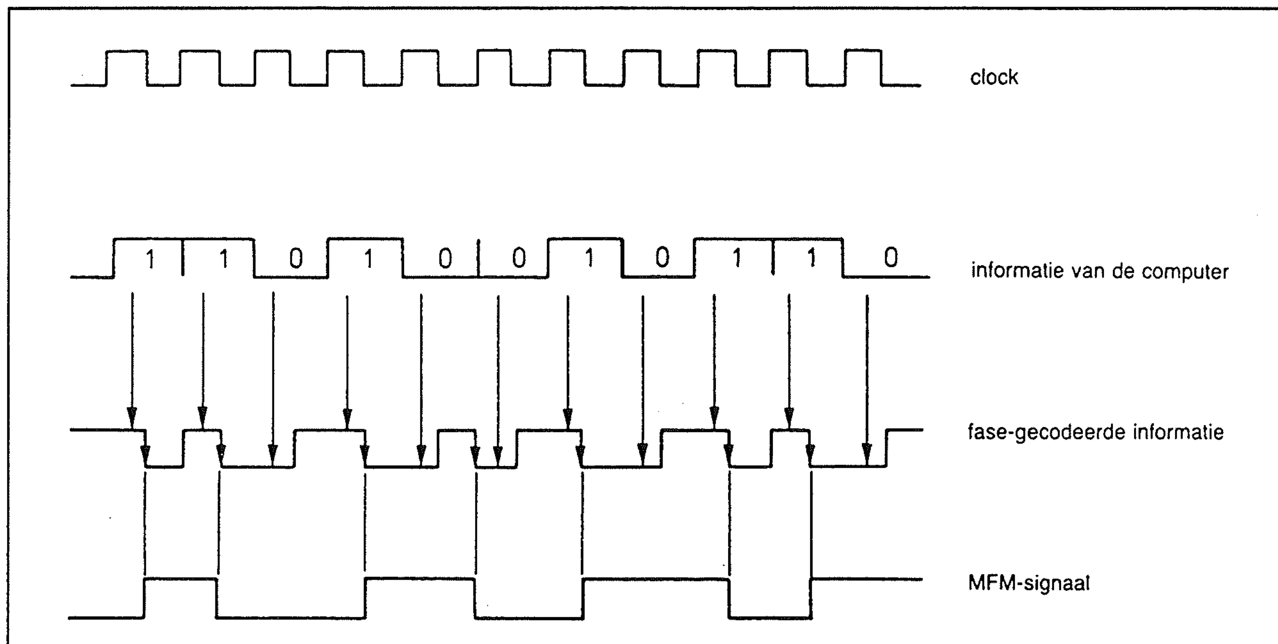
Het FM-systeem

Bij het "Frequentie Modulatie" systeem, kortweg FM genoemd, wordt volgens figuur 3/19.4-12 gebruik gemaakt van een klokpuls met een periode van 4 µs. Deze klokpuls is gesynchroniseerd met de parallelle naar seriële omzetting van de informatie, zodat men ervan verzekerd is dat er bij iedere overgang in het signaal van het ene naar het volgende bit een klokpuls aanwezig is.
Als het bit "L" is wordt alleen de klokpuls naar de schrijfkop gestuurd en wordt een klein gebiedje gemagnetiseerd. Bij een "H"-bit wordt er in het midden van de klokperiode een tweede puls naar de kop gestuurd, zodat er een tweede magnetisatie-zône ontstaat.
Bij het lezen van de schijf kan men de twee signalen scheiden door gebruik te maken van een phase locked loop (PLL). Een oscillator, opgenomen in een fase-terugkoppeling, wordt gesynchroniseerd door de klokpulsen die aanwezig zijn in de eerder besproken "gaten" in een sector. De fase-terugkoppeling zorgt ervoor dat de frequentie van de oscillator precies gelijk wordt aan de frequentie van de van de schijf ontvangen klokpulsen. Een "H"-bit kan dan gedetecteerd worden in een soort venster-discriminator, die onderzoekt of er in de pauze tussen twee klokpulsen al dan niet een puls verschijnt.
Het nadeel van het FM-systeem is dat men er slechts 128 bytes per sector mee kan verwerken.

Het MFM-systeem

Bij dit zogenaamde "Modified Frequency Modulation" systeem wordt een hogere opslagcapaciteit verkregen doordat men de in feite overbodige informatie van de klokpulsen niet op de schijf zit. Het principe wordt toegelicht aan de hand van de grafieken in figuur 3/19.4-13. MFM werkt met een klokperiode van slechts 2 µs. Het kloksignaal wordt in een fasecomparator vergeleken met de gesynchroniseerde seriële datastroom die door de interface wordt aangeleverd.

19.4 Principes van gegevensopslag op diskette



Figuur 3/19.4-13: Het op de schijf opnemen van gegevens volgens het MFM-systeem.

De uitgang van de fasecomparator levert een signaal dat in fase is met de klok als de datastroom "H" is en in tegenfase is met de klok als de datastroom "L" is. Dat heeft tot gevolg dat de richting van de flanken van het uitgangssignaal gelijk is aan de richting van de flanken van het kloksignaal bij hoge informatie en tegengesteld bij lage informatie. Dit fasegemoduleerde signaal stuurt de spanning naar stroomomzetter die de stroompulsen voor het opwekken van het magnetisch veld rond de kopspoel genereert. Bij het terug lezen van de informatie van de schijf wordt weer gebruik gemaakt van een PLL-systeem.
Een oscillator wordt fasegekoppeld aan de referentiebits in de "gaten" op de schijf.
Het van de schijf herwonnen elektrische signaal wordt in een fasevergelijker met dit oscillatorsignaal vergeleken. Deze schakeling vergelijkt de richting van de achterflanken van de twee signalen en kan op deze vrij eenvoudige manier besluiten of er een hoog dan wel een laag signaal gegenereerd moet worden.

Softwarematige organisatie van de schijf

Inleiding
Bij het formatteren van een diskette wordt het volledige bruikbare oppervlak van de schijf ingedeeld in vier componenten, die duidelijk omschreven functies hebben. Deze componenten zijn:
- de boot-record;
- de file allocation tabel (FAT);
- de directory of inhoud;
- de ruimte voor de gegevens.

Boot-record
De boot-record neemt één sector in beslag en wordt ondergebracht in sector 1 van track 0 op zijde 0. Deze file bevat een kort programma dat wordt gebruikt om

19.4 Principes van gegevensopslag op diskette

het eventueel op de schijf aanwezige besturingssysteem in het geheugen te laden. Noteer dat deze file bij iedere schijf ontstaat bij het formatteren, ook al bevat deze schijf alleen gegevensbestanden! Belangrijker in dit kader is echter dat de bootrecord een aantal gegevens bevat over de manier waarop de ruimte op de schijf is ingedeeld. Door het lezen van deze informatie kan het systeem de schijf op de juiste manier benaderen.

File Allocation Tabel

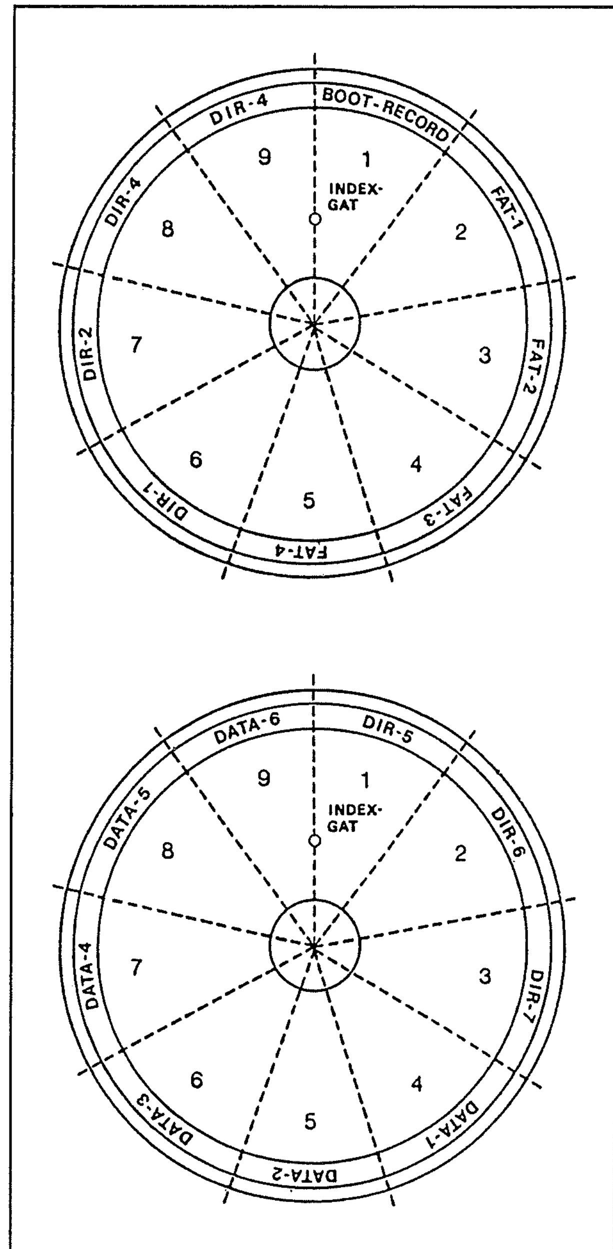
Deze file, die kortweg FAT wordt genoemd, start op sector 2 van track 0 op zijde 0. Deze file staat dus vlak achter de boot-record en bevat de gegevens waaruit het systeem kan afleiden hoe de schijf georganiseerd is. De FAT kan beschouwd worden als de boekhouder van de schijf die noteert:

- hoeveel schijfruimte er in gebruik is;
- hoeveel ruimte er nog vrij is;
- hoeveel ruimte er onbruikbaar is door beschadigingen op het oppervlak van de schijf;
- waar alle files opgeslagen zijn.

Omdat het systeem absoluut niet in staat is de schijf te besturen als er een fout in de FAT aanwezig zou zijn, wordt deze file minstens twee maal op de schijf gezet. De lengte van de FAT is afhankelijk van het soort diskette dat gebruikt wordt en kan variëren van 1 tot 7 sectoren.

Directory

De directory-file bevat de eigenlijke inhoudsopgave van de volledige schijf. Aan de hand van de gegevens in deze file kan het systeem de kop erg snel naar de eerste sector sturen waar een bepaalde file is opgeslagen. De lengte van de directory is uiteraard afhankelijk van de capaciteit van de diskette.



Figuur 3/19.4-14: De softwarematige organisatie van een diskette.

Bij dubbelzijdige diskettes neemt de directory zeven sectoren in beslag, die op de eerste track van beide zijden worden ondergebracht.

Data

De rest van de schijf staat ter beschikking voor het opslaan van gegevens. De secto-

ren in deze ruimte worden in clusters verdeeld die op een basis van "vraag en aanbod" aan de opeenvolgende geschreven files worden toegewezen.

Samenvatting
In figuur 3/19.4-14 is een grafisch overzicht gegeven van de manier waarop de vier besproken basiscomponenten op een dubbelzijdige schijf worden opgeslagen.

De Boot-record

Inleiding
De zogenaamde disk-parameters worden vanaf het vierde byte in de boot-record opgeslagen. Dit 27 byte lange veld bevat alle gegevens die noodzakelijk zijn voor het identificeren van het soort schijf en van de manier waarop deze schijf werd geformatteerd. De tabel van figuur 3/19.4-15 geeft een overzicht van de samenstelling van dit gegevensveld.

START-BYTE	AANTAL BYTES	BETEKENIS
3	8	SYSTEEM-IDENTIFICATIE
11	2	AANTAL BYTES PER SECTOR
13	1	AANTAL SECTOREN PER CLUSTER
14	2	AANTAL GERESERVEERDE SECTOREN VOOR FAT, DIR EN BOOT
16	1	AANTAL FAT'S
17	2	AANTAL DIRECTORY ENTRIES
19	2	AANTAL SECTOREN
21	1	CODE VOOR SOORT SCHIJF
22	2	AANTAL SECTOREN VOOR FAT
24	2	AANTAL SECTOREN PER TRACK
26	2	AANTAL ZIJDEN
27	2	AANTAL GERESERVEERDE SECTOREN VOOR SPECIALE TOEPASSINGEN

Figuur 3/19.4-15: De samenstelling van de gegevens in de boot-record.

De betekenis van de gegevens wordt nu in het kort besproken.

- Systeemidentificatie
  MS-DOS heeft de mogelijkheid ingebouwd dat de fabrikant van het systeem zijn naam en eventueel een typenummer op de schijf zet. Daarvoor staan 8 bytes ter beschikking, deze worden ASCII-gecodeerd.
- Aantal bytes per sector
  Deze bytes bevatten onder HEX-code het aantal bytes dat per sector op de schijf wordt geschreven.
- Aantal sectoren per cluster
  De door de gebruiker/ster op de diskette te zetten gegevens, de eigenlijke gebruikersfiles, worden in de vrije sectoren van de schijf geschreven. Deze sectoren worden ingedeeld in clusters en deze clusters kunnen een of meerdere sectoren lang zijn. In dit byte wordt dit aantal vastgelegd, bijvoorbeeld 01 of 02.
- Aantal gereserveerde sectoren voor FAT, DIR en BOOT
  Zoals uit figuur 3/19.4-14 blijkt, bevat iedere diskette één track op de rand van de schijf die gereserveerd is voor het opnemen van de gegevens van de FAT, de directory en de bootrecord. Deze twee bytes bevatten dus het getal 01 voor zachte schijven, maar omdat dezelfde structuur ook gebruikt wordt voor harde schijven en de situatie er dan heel anders uitziet is het toch noodzakelijk dit voor zachte schijven constante getal te registreren.
- Aantal FAT's
  Bij de meeste diskettes worden er twee kopieën van de file allocation tabel op de schijf gezet.
  Omdat dit aantal soms echter afwijkt van de regel wordt het reëel aantal aanwezige FAT-files in dit byte geregistreerd.
- Aantal directory entry's

In deze twee bytes wordt aangegeven hoeveel entry's er in de directory opgenomen kunnen worden. Dit getal is beperkt en bijgevolg kan het dus voorkomen dat een schijf, waarop men honderden kleine files wil registreren, aangeeft dat zij vol zit, terwijl er volgens de DIR nog honderden kB beschikbaar zijn.

- Aantal sectoren
  Twee bytes die gebruikt worden voor het registreren van het aantal sectoren dat op de schijf aanwezig is, bijvoorbeeld 720 voor dubbelzijdige schijven met negen sectoren per track.
- Code voor soort schijf
  Alle soorten schijven hebben bepaalde HEX-codes gekregen. Dit hexadecimale getal wordt in dit byte geregistreerd.
- Aantal sectoren voor FAT
  In dit twee bit brede woord wordt geregistreerd hoeveel sectoren de file allocation tabel in beslag neemt.
- Aantal sectoren per track
  Uiteraard staat hierin het cijfer 8, 9, 15, 18 of 36.
- Aantal zijden
  Geeft aan of er een enkel- of tweezijdige diskette in de drive zit.
- Aantal gereserveerde sectoren voor speciale toepassingen
  Voor speciale toepassingen kan het noodzakelijk zijn een aantal sectoren te reserveren voor bepaalde files. Dit aantal kan in deze twee bytes worden geregistreerd.

De File Allocation Tabel

Inleiding

Zoals reeds geschreven worden de vrije sectoren van de diskette ingedeeld in een aantal clusters. De naar de schijf te schrijven files worden clustergewijs op de schijf gezet. Wil het systeem een geregistreerde file nadien weer kunnen lezen, dan is het noodzakelijk dat ergens een tabel op de diskette staat waarin vermeld wordt welke clusters aan welke files werden toegewezen. Deze organisatietaak wordt in handen gelegd van de file allocation tabel. Er is echter meer! Het is helemaal niet zeker dat een bepaalde file in aaneengesloten clusters wordt geschreven. Het kan voorkomen dat een file wordt geregistreerd in diverse groepen clusters, die kris-kras over de diskette verspreid liggen. Over het hoe en waarom van deze zogenaamde gefragmenteerde opslag wordt dieper ingegaan in een van de volgende paragrafen. Het zal dus duidelijk zijn dat er een koppeling tot stand moet komen tussen iedere cluster op de schijf en de inhoud daarvan! Ook in de directory van de schijf staan gegevens, de zogenaamde FAT-entry's, die iets met deze tabel te maken hebben. Dank zij deze entry's weet het systeem naar welke positie in de organisatietabel gesprongen moet worden om het begin van een file te vinden. De lengte van de FAT is afhankelijk van het soort schijf, want het zal wel duidelijk zijn dat de tabel grote wordt als de diskette meer gegevens kan bevatten.

De FAT-tabel

De inhoud van de FAT kan beschouwd worden als een tabel, waarin aan iedere cluster van de schijf een kolom is toegewezen.

De clusters worden op een bepaalde manier genummerd van 2 tot en met maximaal 2.372. De inhoud van iedere kolom geeft informatie over het gebruik van de betreffende cluster. Een overzicht:

- Inhoud HEX 000:

De cluster is vrij en kan gebruikt worden voor het opslaan van een nieuwe file.
- Inhoud HEX FF7:
  Bij het formatteren is ontdekt dat er een beschadiging zit op het oppervlak van de diskette waarvan deze cluster gebruik maakt. Deze clusters zullen door het systeem niet gebruikt worden.
- Inhoud HEX FFF:
  Deze cluster is de laatste die door een bepaalde file gebruikt wordt. Na het lezen van deze cluster is de volledige file dus gelezen.
- Overige inhoud:
  Verwijst naar het nummer van de cluster waarin het vervolg van de file is opgeslagen.

Iedere file is dus onder de vorm van een keten van getallen in de FAT-tabel vertegenwoordigd. Bij het lezen of schrijven van een bepaalde file wordt, dank zij de FAT-entry in de directory, naar de cluster in de tabel gesprongen waarin het begin van de file zit. De inhoud van deze kolom verwijst naar het nummer van de volgende cluster waarin het vervolg van de file zit. Zo wordt cluster na cluster de verdeling van de file over de schijf ontrafeld, totdat een kolom wordt gevonden met inhoud HEX FFF. Het systeem weet dan dat dit de laatste cluster is die door de file wordt gebruikt en stopt het leesproces. Een dergelijke ketenvormige katalogering van een file in de FAT-tabel wordt een "allocation chain" genoemd. Iedere file bezit dus een eigen allocation chain in de tabel die, dat zal duidelijk zijn, verweven kan zitten tussen de overige chain's.

Wissen van files
Het zal iedereen wel al eens zijn opgevallen dat het wissen van een file van een diskette veel sneller gaat dan het schrijven van een file met dezelfde omvang. Dat is logisch, want in feite wordt de file niet gewist! Bij het wissen van een file worden alleen alle kolommen van de FAT-tabel die iets te maken hebben met de allocation chain van deze file op HEX 000 gezet. Het systeem weet dan dat deze clusters weer ter beschikking staan voor het opnemen van nieuwe files. De inhoud van deze clusters wordt echter eerst bij het hernieuwd toewijzen van deze clusters aan een nieuwe file overschreven!

De directory

Inleiding
In de directory worden alle gegevens over de afzonderlijke files opgeslagen die van belang zijn voor het snel en efficiënt lezen en schrijven van de diskette. De directory bevat voor iedere file de volgende gegevens:
- de naam;
- de extensie;
- het file-attribuut;
- de FAT-entry;
- de lengte van de file;
- de tijd waarop de file op de schijf werd gezet;
- de datum waarop de file op de schijf is gezet.

Entry's
De directory bevat een aantal zogenaamde "entry's". Iedere entry bevat de gegevens van één file en is 32 bytes lang. Iedere sector van de directory kan bijgevolg 16 entry's bevatten. Omdat er bij een dubbelzijdige schijf 7 sectoren voor de directory gereserveerd zijn, kan zo's schijf maximaal 112 entry's bevatten. Noteer dat er ook voor iedere sub-directory een entry

noodzakelijk is en dat daarnaast ook de gegevens van de naam van de schijf, het zogenaamde "volume label" in een entry worden genoteerd. Een overzicht van de manier waarop al deze gegevens van een file in een entry van de directory worden opgeslagen is gegeven in de tabel van figuur 3/19.4-16.

START-BYTE	INHOUD	AANTAL BYTES	CODE-FORMAAT
0	NAAM	8	ASCII
8	EXTENSIE	3	ASCII
11	ATTRIBUUT	1	BIT
12	VRIJ	10	NULLEN
22	TIJD	2	WOORD
24	DATUM	2	WOORD
27	FAT-ENTRY	2	WOORD
28	LENGTE	4	GETAL

Figuur 3/19.4-16: De samenstelling van een entry van de directory.

File-namen

Het is bekend dat de naam van de file in MS-DOS maximaal acht karakters kan bevatten. Deze worden met hun ASCII-code in de entry geregistreerd. Kortere namen worden aangevuld met spaties, dus CHR$(32). De namen kunnen bestaan uit alle letters van het alfabet, alle cijfers en de tekens / ! @ # $ % & ( ) - _ { } ' ". Spaties zijn verboden, omdat MS-DOS deze niet herkent. Punten zijn natuurlijk ook taboe, omdat deze alleen gebruikt mogen worden om de scheiding tussen de acht karakters van de naam en de drie karakters van de extensie aan te geven. De letters worden als hoofdletters ingevoerd. Van de ASCII-notatie wordt in drie gevallen afgeweken. Het systeem kan in het eerste byte van de naam namelijk enige hexadecimale codes plaatsen. Deze hebben onderstaande betekenis.

- HEX 00:
  Duidt aan dat de betreffende entry niet meer gebruikt wordt. Dit versnelt het doorzoeken van de directory aanzienlijk.
- HEX E5:
  Maakt aan het systeem duidelijk dat de file die in deze entry was opgeslagen ondertussen gewist is van de schijf. Zoals bekend wist het systeem na een DEL-opdracht niet de volledige file van de schijf, maar geeft dit aan in de directory en in de FAT.
  Alleen het eerste byte van de filenaam in de entry wordt overschreven met HEX E5, de rest van de entry blijft ongewijzigd. Vandaar dat het mogelijk is per ongeluk gewiste files onder bepaalde omstandigheden te herstellen. Er bestaan bepaalde programma's die entry's die beginnen met de HEX-code E5 uit de directory selecteren en de rest van de inhoud op het scherm zetten. Het volstaat dan deze code te vervangen door het eerste karakter van de filenaam om de zogenaamd van de schijf gewiste file weer toegankelijk te maken.
  Uiteraard werkt dit systeem alleen als de entry inmiddels niet is toegewezen aan een nieuwe file!
- HEX 2E:
  Is de code die het systeem duidelijk maakt dat de betreffende entry geen file bevat, maar de naam van een subdirectory.

De file-extensie

Deze drie karakters lange code wordt onder de vorm van de ASCII-codes op de schijf gezet. Ook nu worden automatisch spaties ingevuld als men een kortere extensie intoetst of helemaal geen extensie gebruikt.

19.4 Principes van gegevensopslag op diskette

Het file-attribuut

Het ene byte van de attribuut bevat gegevens over het soort file dat in de beschouwde entry wordt gekarakteriseerd. De acht individuele bits van dit byte hebben allemaal een specifieke betekenis, die in de tabel van figuur 3/19.4-17 wordt toegelicht.

- Bit 0:
  Dit bit wordt "H" als de betreffende file alleen gelezen mag worden. Op deze manier kan men files die onder geen enkele omstandigheid veranderd of gewist mogen worden op de schijf zetten.
- Bit 1:
  Verborgen of "hidden" files worden niet herkend door de MS-DOS instructies zoals DIR of COPY. Door dit bit "H" te maken kan men dus een "onzichtbare" file op de schijf zetten.
- Bit 2:
  Ook systeemfiles zijn verborgen en in feite bestaat er geen wezenlijk verschil, althans wat betreft schijf-systematiek, tussen een verborgen en een systeem file.
- Bit 3:
  Dit bit wordt "H" als de entry niet de gegevens van een file bevat maar wordt gebruikt voor het onderbrengen van de naam van de diskette. Dit onderscheid is noodzakelijk omdat diskettenamen elf karakters lang kunnen zijn. Deze naam wordt als ASCII-code opgeslagen in de bytes die normaal voorbehouden zijn voor de filenaam en -extensie.
- Bit 4:
  Wordt "H" als de entry de naam van een sub-directory bevat.
- Bit 5:
  Deze back-up vlag heeft alleen betekenis bij harde schijven. Daar wordt door het "H" worden van dit bit aangegeven dat de file is geactualiseerd. Op deze manier kan het systeem heel snel alleen deze files waar iets in veranderd is backuppen naar een schijf. Bij diskettes is dit bit steeds "L".

BIT WORDT "H" VOOR	BETEKENIS
0 LSB	READ-ONLY FILE
1	VERBORGEN FILE
2	SYSTEEM-FILE
3	VOLUME LABEL
4	SUB-DIRECTORY
5	BACK-UP VLAG
6	ONGEBRUIKT
7 MSB	ONGEBRUIKT

Figuur 3/19.4-17: De betekenis van de acht bits uit het attribuut-byte.

De tijd- en datumregistratie

De bytes 22 en 23 bevatten de tijd waarop de file voor het laatst naar de schijf werd geschreven onder de vorm van een geheel getal. De waarde van dit getal wordt gegeven door de uitdrukking:
GETAL = (2048 x uren) + (32 x minuten) + (0,5 x seconden).
De datum wordt in de bytes 24 en 25 van de entry genoteerd onder de vorm van een geheel getal waarvan de waarde wordt gegeven door de uitdrukking:
GETAL = (512 x (jaar - 1980)) + (32 x maand) + (dag).

De FAT-entry

In dit twee bytes lange woord wordt genoteerd in welk cluster van de diskette het begin van de file wordt opgenomen. Dit woord verwijst dus naar de kolom in de FAT-tabel waarin de betreffende file voor het eerst genoteerd wordt. Op deze ma-

nier wordt de kop van het leesmechanisme snel naar het juiste track verplaatst.

De file-lengte

Vier bytes staan ter beschikking voor het noteren van de karakterlengte van de file. Dat wil zeggen dat alleen de fysische capaciteit van de schijf een grens stelt aan de maximale file-lengte! In deze vier bytes kan immers een getal opgenomen worden dat veel groter is dan de capaciteit van de schijf.

De ruimte voor gegevens

Inleiding

Het grootste gedeelte van de diskette staat ter beschikking voor het opslaan van gegevens en/of programma's. Zoals reeds geschreven is deze ruimte verdeeld in clusters, die ieder minstens twee sectoren bevatten. Bij een nieuw geformatteerde schijf wordt de eerste file opgeslagen in de eerste clusters van het gegevensgebied. Als deze file opnieuw naar de schijf wordt geschreven en het systeem merkt dat de file langer is geworden, dan worden er nieuwe clusters ter beschikking gesteld. Bij het kopiëren van een aantal bestanden naar een lege diskette worden alle files dus in opeenvolgende clusters opgeslagen.

Gefragmenteerde opslag

Heel anders gaat het er echter aan toe als een nieuwe file wordt opgeslagen op een diskette die al lang in gebruik is en waar al vaak files van gewist zijn. Er zijn dan "gaten" in het datagebied en deze gaten worden gebruikt voor het schrijven van nieuwe files. Het kan dan uiteraard gebeuren dat zo'n "gat" te klein is voor een lange file. De file wordt dan verdeeld over verschillende, vaak ver uit elkaar liggende tracks van de schijf. Dit noemt men het gefragmenteerd opslaan van gegevens. Dit heeft als groot voordeel dat de capaciteit van de schijf onder alle omstandigheden gelijk blijft. Groot nadeel is echter dat het systeem de kop heen en weer moet bewegen over de schijf om de juiste posities van alle clusters op te zoeken waarin delen van de file zijn opgeslagen. Dit kan het inlezen van een file aanzienlijk vertragen. Als men dit bezwaarlijk vindt bestaat er een eenvoudige oplossing voor dit probleem. Als men namelijk file na file van de gefragmenteerde schijf kopieert naar een nieuwe schijf, dan worden alle files weer netjes in opeenvolgende clusters opgenomen.

De functie van de FAT

Zoals reeds geschreven bij de bespreking van de FAT wordt de verdeling van een file over verspreide clusters genoteerd in de file allocation tabel. Door het inlezen van deze gegevens weet het systeem welke clusters achtereenvolgens uitgelezen moeten worden om de volledige file in de juiste volgorde uit te lezen.

3/19.5

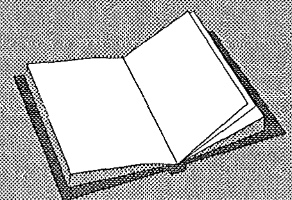
Principes van gegevensopslag op harde schijven

Inleiding

Geschiedenis
De technologie van de harde schijf heeft een stormachtige ontwikkeling doorgemaakt. In de jaren zestig bracht IBM de eerste harde schijf op de markt. Het kastgrote apparaat had een opslagcapaciteit van 30 MB, terwijl er nog eens 30 MB beschikbaar was onder de vorm van verwisselbare schijven. Het apparaat leek dan ook meer op een jukebox dan op een moderne harde schijf! Dit apparaat ging al gauw door het leven onder de naam "30-30". In die tijd werden er echter nogal wat westernfilm's vertoond met John Wayne in de hoofdrol, die met zijn beruchte Winchester 30-30 onverbiddelijk alle misdadigers uit deze wereld hielp. Binnen de kortste keren werd het indrukwekkende apparaat van IBM dan ook "Winchester" genoemd! Deze bijnaam voor een harde schijf is vrij lang in gebruik geweest, ook nadat harde schijven veel meer capaciteit hadden dan de 30+30 MB van IBM's uitvinding. In 1980 introduceerde Seagate haar ST506 loopwerk. Uitgerust met 5.1/4 inch schijven met een capaciteit van slechts 5 MB, was dit apparaat groter dan een moderne computer. Tegenwoordig worden er harde schijven aangeboden met een capaciteit van 2.000 MB, die zijn uitgerust met miniatuur 3.1/2 inch schijven en zo klein zijn, dat zij gemakkelijk in een draagbare computer kunnen worden ingebouwd.

De naamgeving
Over de naamgeving van de harde schijf bestaan verschillende opvattingen. Soms wordt gesproken van "vaste schijf", omdat het apparaat vast in het computersysteem is ingebouwd en de schijf niet verwijderd kan worden. Maar tegenwoordig zijn er natuurlijk ook harde schijven in omloop, waarvan het loopwerk wél vast in de computer is ingebouwd, maar waarvan men de schijven kan wisselen. Erg steekhoudend is de naam "vaste schijf" dus niet meer. Het begrip "harde schijf" verwijst naar de struktuur van de schijven die gebruikt worden om de gegevens op te slaan. In tegenstelling tot hun soortgenoten bij diskettes, zijn deze schijven namelijk altijd gemaakt van een hard materiaal en zijn dus onbuigzaam. Deze naam voldoet dus

LEES OOK:
Hoofdstuk 6/10.21



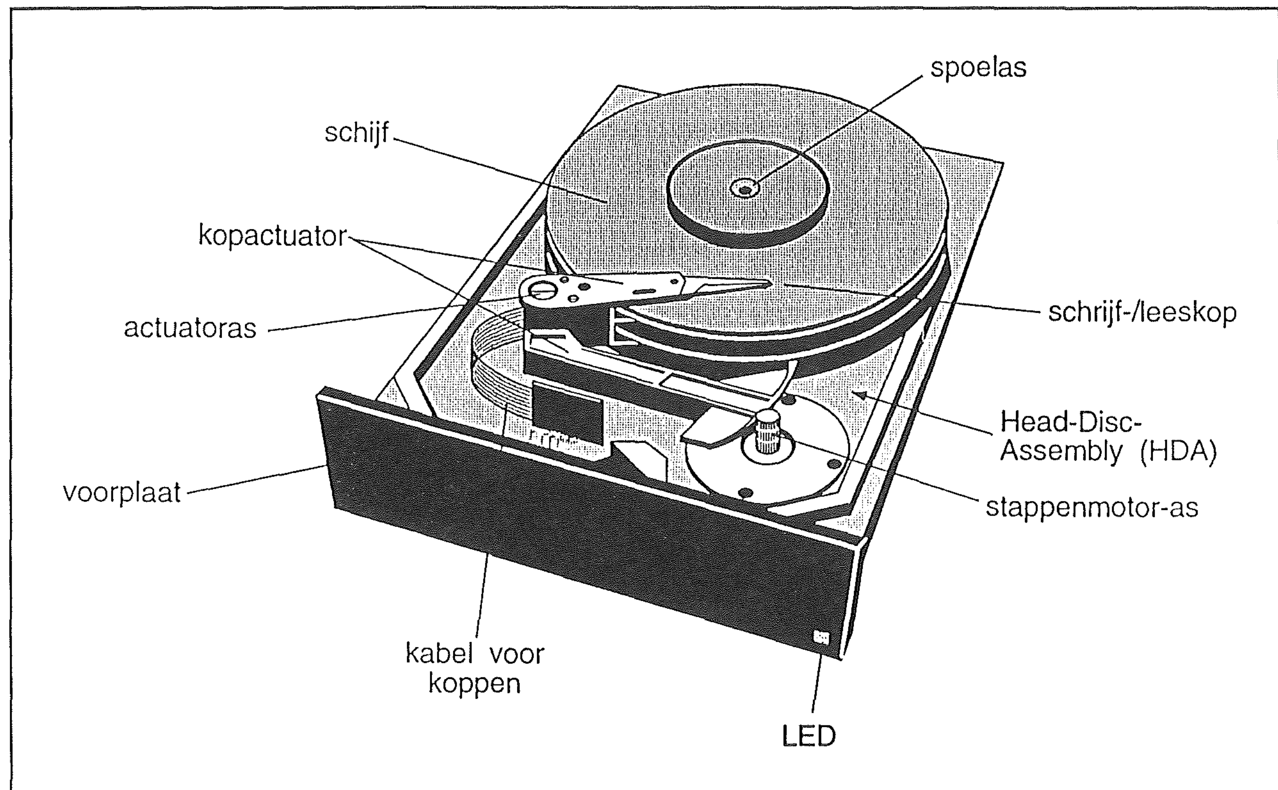
beter en sluit bovendien mooi aan op de Engelse benaming "hard disk", die internationaal is ingeburgerd. Maar om volledig consequent te zijn, zou men diskettes "zachte schijven" moeten noemen, een naam die echter niet erg is ingeburgerd.

De opbouw van een harde schijf

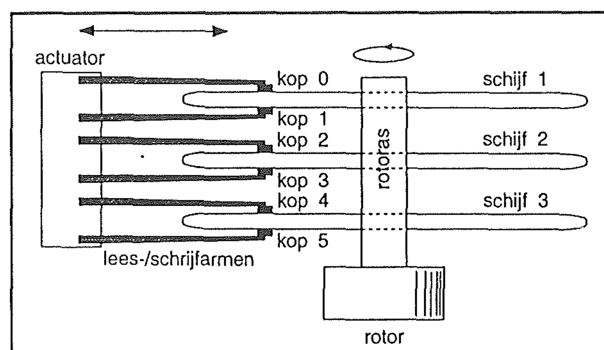
Verschillen tussen harde en zachte schijven
Een harde schijf werkt in theorie net zo als een floppy disk, maar er zijn toch een aantal significante verschillen:
- een floppy disk bestaat uit één schijf die flexibel is, terwijl een harde schijf uit meerdere schijven bestaat die (meestal) een stalen ondergrond hebben en daardoor volledig star zijn;
- bij een floppy disk is de elektronica niet bij de disk gemonteerd en bij de harde schijf zit de elektronica in de HDA, de zogenoemde "Head Disc Assembly";
- bij een floppy disk is er een aandrijfpunt om deze te laten draaien, terwijl er bij een harde schijf gebruik wordt gemaakt van een starre as;
- de behuizing van een harde schijf *moet* luchtdicht worden afgesloten, terwijl dit bij een floppy disk niet noodzakelijk is;
- een harde schijf kan veel meer informatie opslaan dan een floppy disk, namelijk gemiddeld 30 tot 1.500 keer;
- een harde schijf draait gemiddeld 12 keer zo snel als een floppy disk:
  harde schijf: 3.600 rpm;
  floppy disk: 300 rpm.
  3.600 rpm komt overeen met een omwentelingstijd van 17 ms.

Fysieke opbouw van een harde schijf
De fysieke opbouw van de harde schijf is samengevat in figuur 3/19.5-1, terwijl het basisprincipe van de werking beschreven is in figuur 3/19.5-2.
Het volledig mechanisme van de harde schijf is ondergebracht in de reeds genoemde HDA.
De afbeelding van figuur 3/19.5-1 gunt de lezer(es) een kijkje in het inwendige van deze volledig luchtdicht afgesloten behuizing.
De volgende onderdelen vallen op:
- Schijven
  De informatiedrager van de harde schijf.
- Spoelas
  Het aandrijfpunt voor de schijven.
- Schrijf/leeskop
  Deze worden gebruikt om informatie op de schijven op te slaan of informatie van de schijven te lezen.
- Kop-actuator
  Dient ervoor om de registratie op de goede positie van de schijf te plaatsen.
- Actuator-as en stappenmotor-as
  Worden gebruikt voor de aansturing van de arm.
  Voor harde schijven van de eerste generaties werd gebruik gemaakt van een stappenmotor. In moderne harde schijven wordt de kop-actuator door middel van een elektro-magnetisch systeem met spreekspoel heen en weer bewogen, net zoals de LASER bij een CD-speler.
- Elektronica (niet getekend)
  Heeft als functie om de harde schijf correct te laten functioneren. Dit geldt zowel voor de motor als voor de aansturing.
- Behuizing
  De behuizing biedt bescherming aan de eerder beschreven onderdelen.

19.5 Principes van gegevensopslag op harde schijven



Figuur 3/19.5-1: Een inkijkje in de HDA, de "Head Disk Assembly" waarin alle mechanische componenten van de harde schijf aanwezig zijn.



Figuur 3/19.5-2: Het principe van een harde schijf met drie starre schijven.

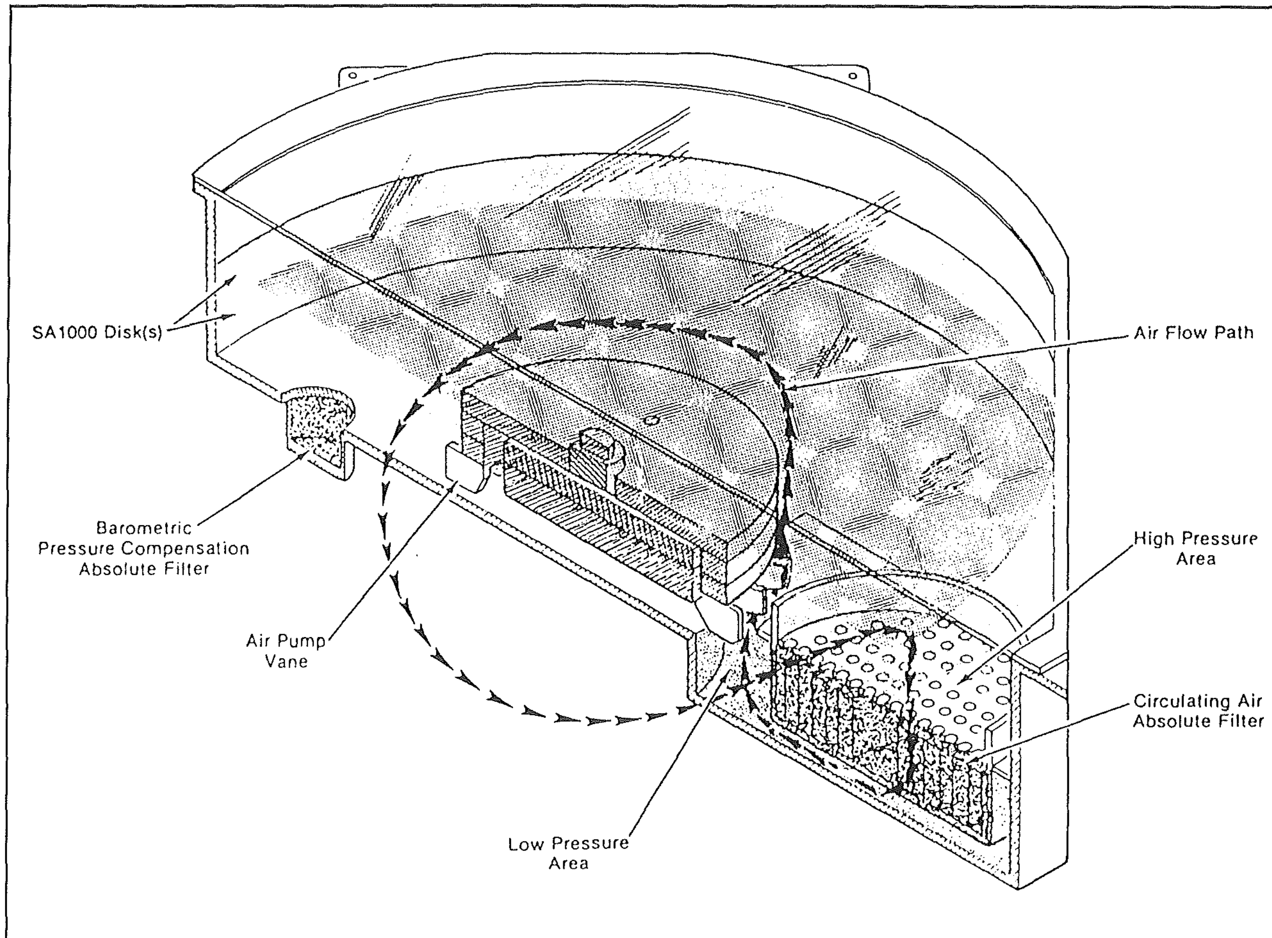
Basisprincipe

Figuur 3/19.5-2 geeft het basisprincipe van de fysieke opbouw van de harde schijf weer. Hier ziet men een viertal schijven die gemonteerd zijn op een vaste as, zodat alle schijven met dezelfde snelheid draaien. Bij elke schijf is aan de boven- en onderzijde een lees/schrijfkop (registratiekop) bevestigd, die allemaal zelf weer op één punt verbonden zijn.

Dit gemeenschappelijke punt is de arm van de registratiekoppen, die op zijn beurt aangestuurd wordt door de spreekspoel of de stappenmotor. Eerder is gesteld dat een harde schijf luchtdicht afgesloten is. Dat is noodzakelijk, want de registratiekoppen van de harde schijf staan op een minimale afstand van de schijf zélf. Als er dus geen goede voorzieningen getroffen worden tegen het indringen van stofdeeltjes, loopt men grote risico's op een zogenaamde headcrash.

Dit verschijnsel, waarbij de koppen van de drive contact maken met het oppervlak van de schijf, heeft meestal een gedeeltelijke vernietiging van de gegevens op de harde schijf tot gevolg.

19.5 Principes van gegevensopslag op harde schijven



Figuur 3/19.5-3: De luchtcirculatie in een Shugart SA1000 harde schijf.

Luchtcirculatie van wezenlijk belang

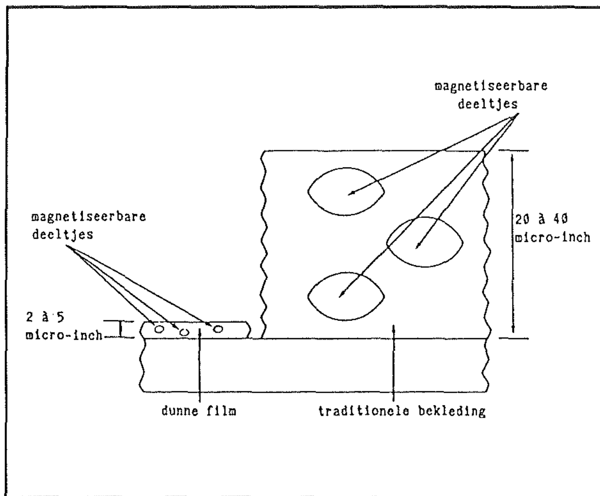
Om een zo efficiënt mogelijke werking van de harde schijf te kunnen handhaven, moet er een zo constant mogelijke luchtcirculatie zijn. De registratiekop zweeft immers boven de harde schijf op een luchtstroom. De luchtcirculatie van een Shugart SA1000 8 inch harde schijf staat getekend in figuur 3/19.5-3.

Alle harde schijven bezitten een luchtfilter, dat in eerste instantie is bedoeld voor het filteren van de lucht in de behuizing. In de meeste gevallen vindt er alleen als de luchtdruk of de temperatuur van de harde schijf verandert een zeer geringe uitwisseling van lucht plaats tussen de behuizing en de buitenwereld. Hoewel het erg vervelend is dat men moet toelaten dat lucht van buiten in de HDA doordringt, is dit absoluut noodzakelijk. Als men een draagbare computer zou meenemen met een vliegtuig en de bagage wordt niet in een drukcabine bewaard, dan zou de interne hoge druk van een volledig afgesloten HDA een kracht van 50 N op de behuizing uitoefenen! De kans is groot dat de harde schijf letterlijk zou exploderen!

Magnetisme en luchtspleet

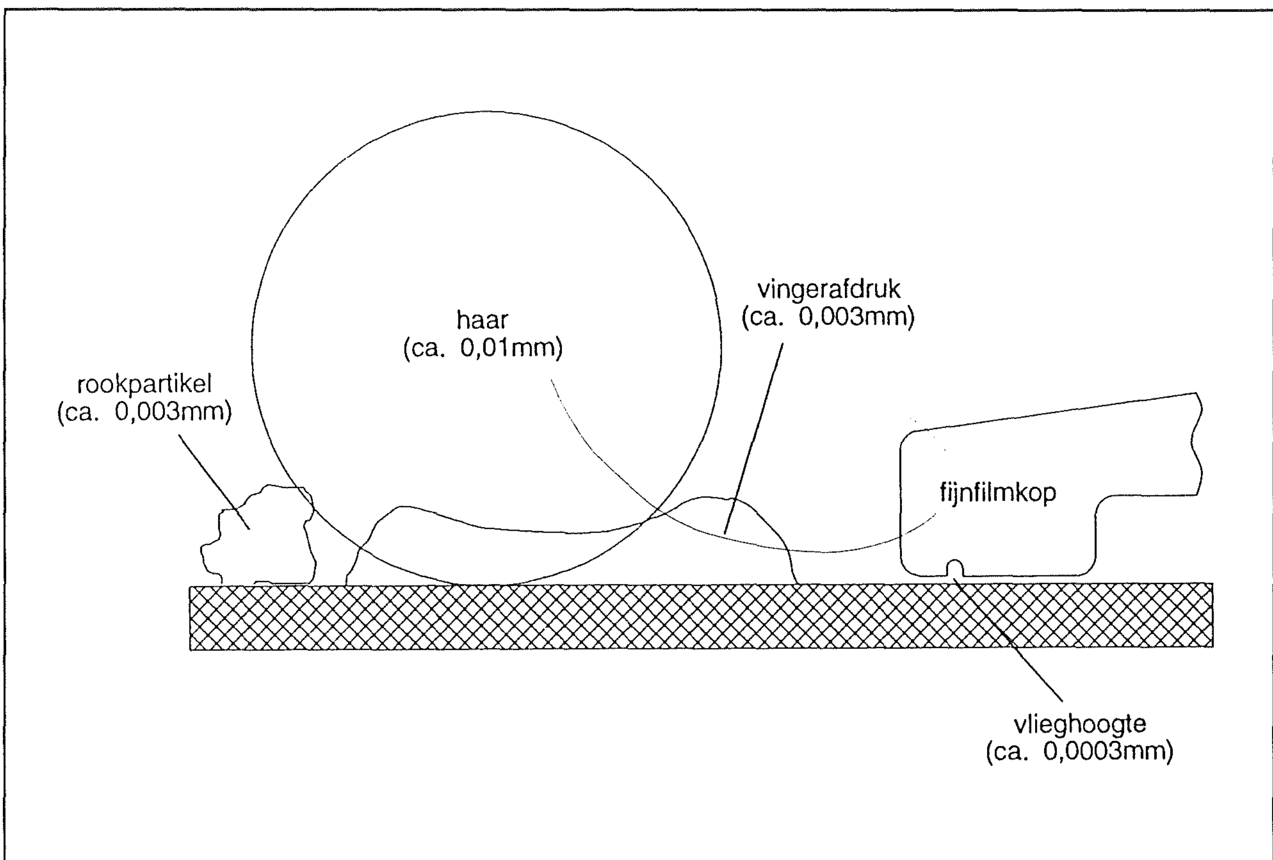
Het belang van een absoluut stofvrije lucht in de HDA volgt uit het volgende. Een harde schijf bestaat uit verschillende stalen schijven, die allen tesamen op een vaste as gemonteerd zijn.

19.5 Principes van gegevensopslag op harde schijven



Figuur 3/19.5-4: Vergelijking van de dikte van de magnetische lagen van een oude en een moderne harde schijf.

Op iedere schijf is aan de bovenkant en de onderkant een magnetische laag opgebracht. Deze laag bestaat in de oudere type's uit een laag metaaloxyde, terwijl in de nieuwere harde schijven gebruik wordt gemaakt van dunne film technologie. De dunne film laag is circa tien maal zo dun als de metaaloxyde laag, zie hiervoor figuur 3/19.5-4. De kop van de harde schijf zweeft op ongeveer 3 μm boven de magnetische laag. Deze afstand is onvoorstelbaar klein en krijgt enig perspectief als zij in figuur 3/19.5-5 wordt vergeleken met de dikte van een haar, een vingerafdruk en een rookdeeltje. Het zal dus duidelijk zijn dat de lucht in het interne van de HDA absoluut zuiver moet zijn!



Figuur 3/19.5-5: De minimale afmetingen van de luchtspleet tussen harde schijf en kop toegelicht aan de hand van een paar voorbeelden.

De harde schijven

Moderne standaard harde schijven zijn uitgerust met schijven die een diameter van 5.1/4 inch of 3.1/2 inch hebben. Voor draagbare computer zijn er echter ook loopwerken met schijven van 2.1/2 inch en zelfs 1 inch in productie. Toch hebben ook deze loopwerken een zeer grote capaciteit.

Kleine schijven hebben een aantal voordelen. Op de eerste plaats zal de toegangstijd tot de gegevens kleiner zijn, omdat de kopbewegingen over het oppervlak van de schijven veel kleiner zijn. Op de tweede plaats maken loopwerken met kleinere schijven veel minder geluid.

Vroeger werden de schijven van ijzer gemaakt. De meeste moderne schijven bestaan echter uit een aluminium-legering, dit vanwege het lage soortelijk gewicht van dit metaal. De schijven zijn dus licht, hetgeen betekent dat de motor veel sneller in staat is de schijven op hun omwentelingssnelheid te brengen. Dit is vooral bij accugevoede computers van groot belang, omdat de motor van de harde schijf hierbij alleen wordt ingeschakeld als het systeem gegevens van de schijf wil lezen of er gegevens naar toe schrijven. Op de aluminium schijven wordt uiteraard een magnetiseerbare laag aangebracht, zoals kobalt of speciale ijzerceramieken. De lagen zijn zeer dun, een paar micron is tegenwoordig standaard. Men spreekt dan ook van dunne laag technologie, in tegenstelling tot de dikke laag technologie die bij de vorige generaties harde schijven werd toegepast. De gevoelige lagen worden ofwel opgedampt, ofwel gesputterd ofwel door middel van de middelpuntvliegende kracht over het oppervlak van de schijf verdeeld. De verschillende technologieën hebben te maken met de dikte van de magnetische laag. Dikke lagen worden middelpuntvliegend aangebracht, dunne lagen door middel van opdamping of sputtering. De dikte van de laag heeft weer te maken met de capaciteit van de harde schijf. Hoe groter de gegevensdichtheid op de schijven moet zijn, hoe minder dik men de lagen moet maken. In de meeste gevallen wordt de magnetische laag bespoten met een uiterst dunne laag harde beschermingslak, waarna het oppervlak volledig glad gepolijst wordt. Iedere oneffenheid, hoe klein ook, kan immers tot gevolg hebben dat de kop in botsing komt met het oppervlak van de schijf.

De werking van een harde schijf

Inleiding

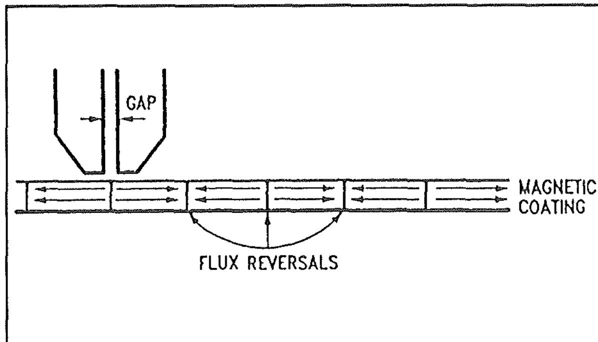
Dit subhoofdstuk behandelt de werking van de harde schijf, niet electronisch bekeken maar fysisch. Aspecten die behandeld worden zijn:

- magnetisme (basis voor de harde schijf);
- registratiekoppen;
- headcrash.

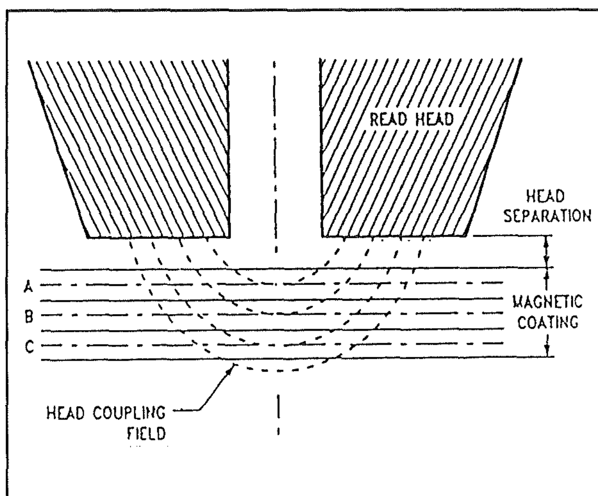
Magnetisme

Magnetisme is de basis van de gegevensopslag op een harde schijf. De enen en nullen van de informatie moeten dus vertaald worden naar magnetische velden, die kleine lokaties op de harde schijf magnetiseren. De nullen en enen hebben hun eigen magnetische polariteit. De magnetische polariteit komt overeen met de richting van de magnetische krachtlijnen van de kleine magneetjes die op de oppervlakte van de harde schijf worden aangebracht.

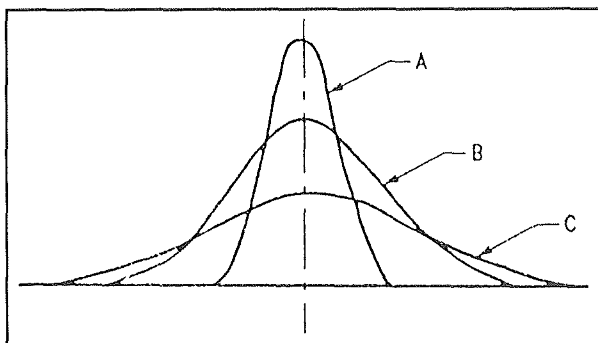
19.5 Principes van gegevensopslag op harde schijven



Figuur 3/19.5-6: De richting van de magnetische krachtlijnen of de polariteit van de gemagnetiseerde locaties op de harde schijf.



Figuur 3/19.5-7: De magnetische krachtlijnen die tussen de twee polen van de schrijfkop ontstaan.



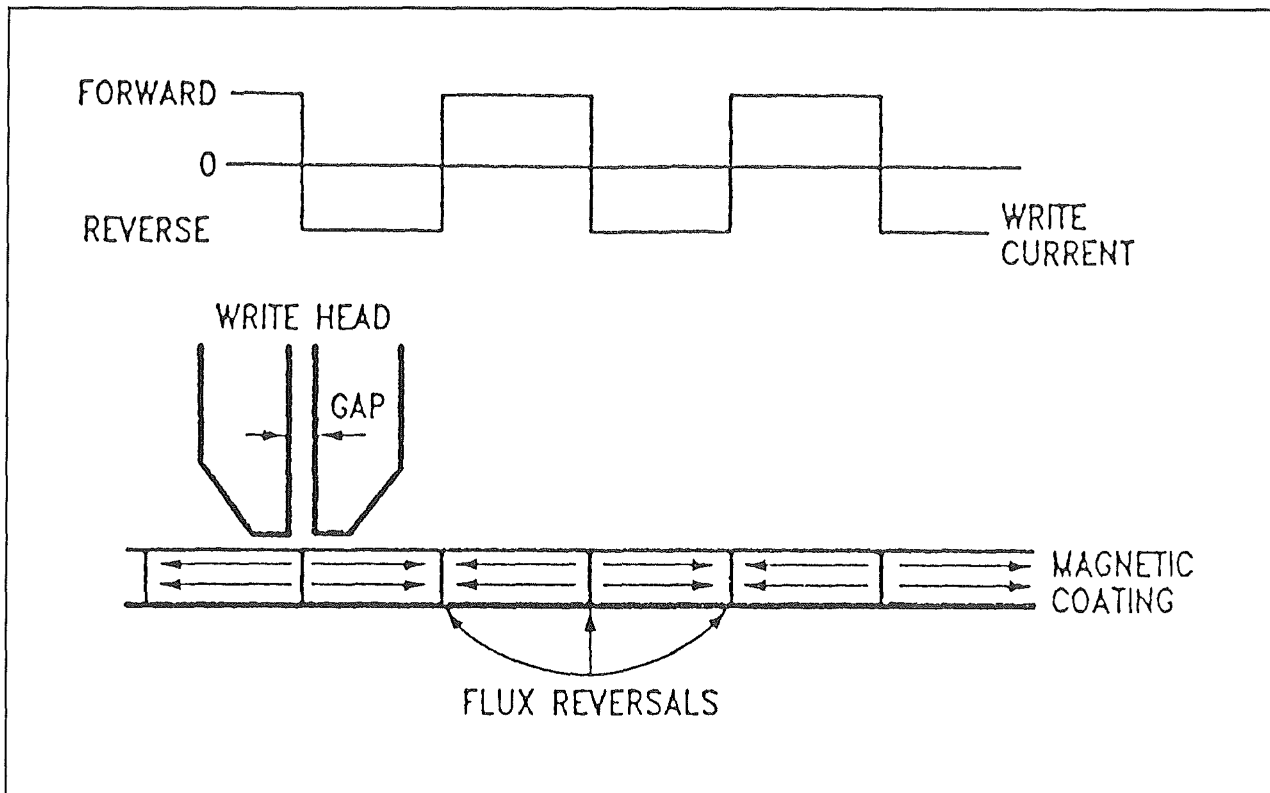
Figuur 3/19.5-8: De verdeling van het magnetisch moment onder de schrijfkop.

Als men uitgaat van een seriële stroom van nullen en enen die op de harde schijf opgeslagen wordt dan betekent dit dat de nul een polariteit naar rechts heeft en de één een polariteit naar links. Dit is voorgesteld in figuur 3/19.5-6.

Schrijven en lezen

Om de data op de harde schijf te krijgen en deze er nadien weer vanaf te kunnen halen, maakt men gebruik van het schrijf en lees principe. Hiervoor wordt een beroep gedaan op een registratiekop (die in detail verderop behandeld wordt), die te vergelijken is met de kop van een cassette recorder. Voor het schrijven op een harde schijf wordt de datastroom toegevoerd aan een spoeltje op de registratiekop. Deze kop wordt daardoor gemagnetiseerd en stuurt vervolgens magnetische krachtlijnen van de ene zijde naar de andere zijde van de registratiekop.
Dit noemt men de twee polen van de kop. Dit principe is weergegeven in figuur 3/19.5-7. De stippellijnen die via een boog van de ene pool naar de andere pool lopen, stellen de magnetische krachtlijnen voor. Deze magnetiseren het gebied van de laag op de schijf dat onder de kop zit. Des te dichter een stippellijn bij de registratiekop is, des te sterker is de krachtlijn. Dit betekent dat gebied A het sterkste magnetische moment heeft en dus het meest betrouwbare is en gebied C het zwakste magnetische moment heeft. Dit verschijnsel staat grafisch weergegeven in figuur 3/19.5-8. De reden waarom het magnetische moment van gebied A veel steiler is dan dat van gebied C is dat gebied C veel meer last van strooivelden heeft dan gebied A. In figuur 3/19.5-9 staat het basis principe van het schrijven op een magnetisch oppervlak, dit hoeft dus niet perse een harde schijf zijn.

19.5 Principes van gegevensopslag op harde schijven



Figuur 3/19.5-9: Het basis principe van het schrijven van gegevens op de harde schijf.

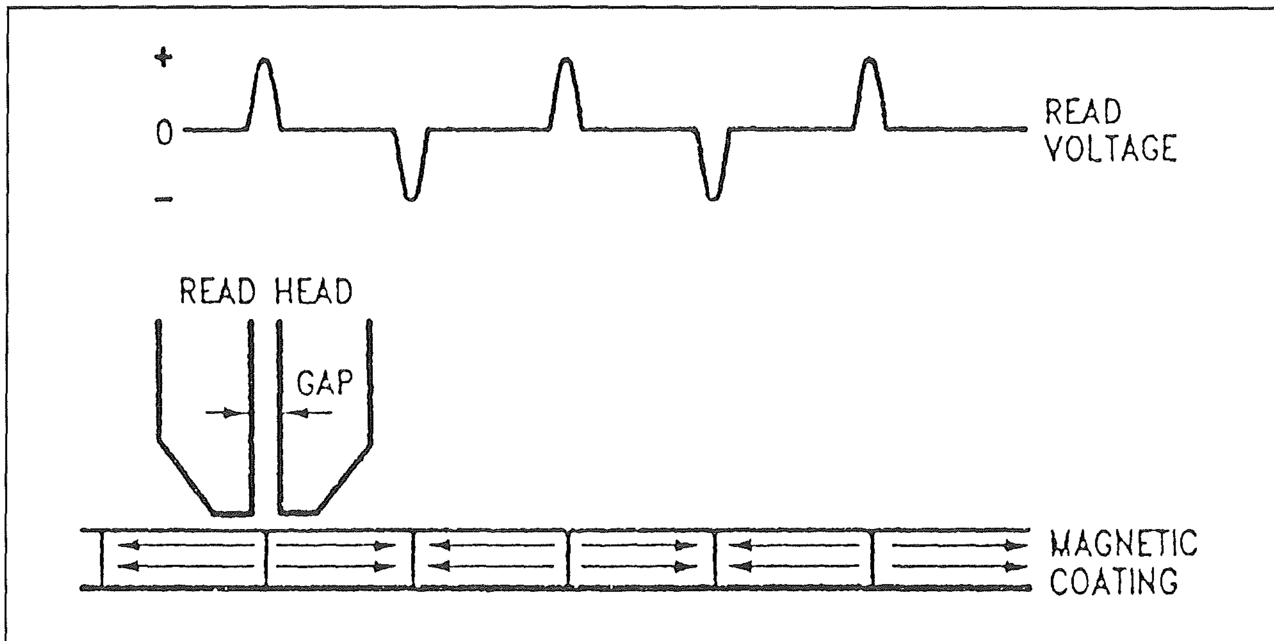
Hiervoor bestaan ook nog andere opslag media, zoals bijvoorbeeld een floppy disk of een cassette bandje. Het bovenste deel van figuur 3/19.5-9 stelt de datastroom naar de registratiekop voor, met daaronder de registratiekop en een magnetisch oppervlak.

Zoals eerder beschreven staat, stellen de nullen een pijl naar rechts voor en de enen een pijl naar links. Dit is de richting van het magnetisch veld. Helemaal onderaan het magnetische oppervlak staat nog een aantal lijnen dat aangeeft waar bijvoorbeeld van richting veranderd wordt, hetgeen gebeurt als een databit verandert.

Voor het lezen van data wordt meestal dezelfde registratiekop gebruikt als voor het schrijven. Deze staat stil boven de snel ronddraaiende schijf.

Op ieder punt van het oppervlak staat een magnetisch moment. De magnetische krachtlijnen rond deze momenten wekken in de kern van de kop een magnetisch veld op. De richting van dit veld is afhankelijk van de richting waarin de momenten op de schijf bij het lezen werden gepolariseerd. De magnetische krachtlijnen in de kern van de kop wekken in de spoel van de kop kleine spanningspulsen op. De polariteit van deze pulsen is afhankelijk van de richting van het magnetisch veld door de kern van de kop. Door de opeenvolgende magnetische momenten op het oppervlak van de schijf op de beschreven manier af te lezen krijgt men een reeks pulsen, waarvan de polariteit aangeeft welk databit er werd uitgelezen, dus een nul of een één. Dit proces is toegelicht in figuur 3/19.5-10.

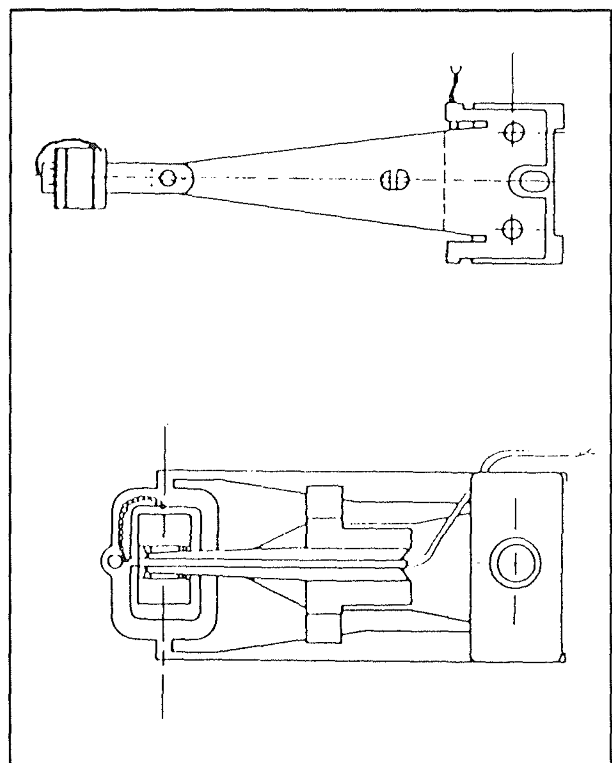
19.5 Principes van gegevensopslag op harde schijven



Figuur 3/19.5-10: Het basis principe van het lezen van gegevens van de harde schijf.

De registratiekop

Zoals beschreven wordt het magnetisch veld door middel van een registratiekop op het magnetische oppervlak van de schijf overgebracht. Afhankelijk van de manier waarop de magnetische laag op de aluminium platen werd aangebracht, besluit de fabrikant tot het toepassen van een van twee technologieën. De eerste registratiekop is de alom bekende Winchester kop die gebruikt wordt voor de oude dikke magnetische laag. Met de ontwikkeling van de dunne film magnetische lagen is er tevens de behoefte ontstaan aan een andere registratiekop, die de Whitney kop wordt genoemd. Beide registratiekoppen worden voorgesteld in figuur 3/19.5-11. Doordat de harde schijf van binnen luchtdicht afgesloten is, is het mogelijk om de registratiekop op een minimale afstand te laten zweven boven het magnetische oppervlak. Deze minimale afstand is variabel van 10 µinch (10^{-6} inch) tot en met 18 µinch.



Figuur 3/19.5-11: De twee gebruikte registratiekoppen, respectievelijk de Winchester en de Whitney.

19.5 Principes van gegevensopslag op harde schijven

Dit komt overeen met 25 miljoenste millimeter tot circa 50 miljoenste millimeter! Het zal dus wel duidelijk zijn dat er absoluut geen verontreinigingen op het oppervlak van de kop aanwezig mogen zijn. Zou dit wel het geval zijn, dan kan men er vrijwel van verzekerd zijn dat de harde schijf op een of meerdere plaatsen grondig vernield zal worden.
De vernieling van een harde schijf ten gevolge van stof of andere vuildeeltjes wordt een "headcrash" genoemd. Een headcrash betekent niets anders dan dat de registratiekop met een grote klap op het magnetisch oppervlak terecht komt, waardoor de magnetische laag en de registratiekop zelf vernield worden. Zoals dus wel duidelijk zal zijn, is zelfs het kleinste vuildeeltje (rook) nog 25 maal zo groot als de opening tussen de registratiekop en het magnetisch oppervlak. Zelfs een rookdeeltje mag dus niet in de harde schijf binnen dringen. Vandaar dat het absoluut noodzakelijk is een harde schijf volledig luchtdicht van de buitenwereld af te sluiten.

Controllers voor de harde schijf

Inleiding
De "harde schijf controller" bevat alle noodzakelijke elektronica voor het besturen, beschrijven en lezen van de harde schijf. Er bestaan vier verschillende standaarden, die in de loop der tijd zijn ontwikkeld.
- SMD interface;
- ST412/ST506 interface;
- SCSI interface;
- ESDI interface.

De volgende (de)coderingstechnieken worden toegepast:
- MFM (de)codering;
- RLL (de)codering;
- NRZ (de)codering.

Deze interfaces en hun (de)codering zijn uitvoerig besproken in hoofdstuk 6/10.21 en worden hier niet opnieuw behandeld.

Fysieke installatie van een harde schijf

Inleiding
Onder de fysieke installatie van een harde schijf wordt verstaan al de handelingen die nodig zijn om ervoor te zorgen dat de gebruiker een harde schijf beschikbaar heeft die alle basishandelingen kan verrichten.
Dit betekent niet:
- dat de harde schijf logisch ingedeeld moet zijn, dit is namelijk de initialisatie en de indeling van de directory structuur;
- dat er programmatuur op de harde schijf aanwezig is.

Dit geldt zowel voor DOS als voor andere programmatuur, want men kan ten allen tijde een computer starten vanaf de floppy disk.

Fysieke installatie
Het fysiek installeren van een harde schijf kan men het best volgens een vast schema doen.
Een geadviseerd schema bevat de volgende zeven stappen.
- Zet de computer uit, verwijder alle stekkers (met uitzondering van het netsnoer) uit de connectoren en open de computer.

19.5 Principes van gegevensopslag op harde schijven

- Raak de buitenkant van de computer aan om de statische lading te verwijderen en verwijder het netsnoer.
- Maak de benodigde ruimte voor de installatie van de harde schijf vrij.
- Maak de benodigde ruimte voor de installatie van de controller kaart vrij.
- Zet alle kabels aan de controller kaart en plaats deze in het gewenste slot. Eventueel kunnen nu ook de jumpers op de controllerkaart ingesteld worden.
- Verander, indien noodzakelijk, de jumpers van de computer die nodig zijn om de harde schijf te herkennen.
- Plaats de harde schijf in de computer, bevestig alle kabels en werk deze netjes weg.

Stap 1
Zet eerst met de schakelaar de computer uit, verwijder nadien alle bedrading aan de achterkant van de computer en draai daarna de schroeven aan de achterzijde los, zodat de systeemkast verwijderd kan worden. In het tweede gedeelte van deze stap dient men alle bekabeling aan de achterkant van de computer te verwijderen. Hier dient men goed in de gaten te houden welke kabel waar zit. Als er veel kabels zijn kan het nuttig zijn om kabels te merken of om een tekening te maken welke kabels waar verbonden zitten.
Vervolgens kan men de systeemkast verwijderen, hierbij dient men echter goed in de gaten te houden dat de stekker van het netsnoer nog in het stopcontact zit. Dit doet men door allereerst de systeemkast naar voren te schuiven totdat deze niet meer verder wil.
Daarna moet men de systeemkast iets terug schuiven, deze oplichten en daarna weer verder naar voren schuiven tot deze verwijderd is.

Stap 2
Deze stap dient ervoor om de aanwezig statische electriciteit van het menselijk lichaam te neutraliseren. Vandaar dat het netsnoer nog niet verwijderd mocht worden. De computer is nu nog geaard via het netsnoer en de statische elektriciteit kan via de systeemkast afvloeien naar de aarde. Statische electriciteit moet verwijderd worden, omdat deze spanningen tot 4.000 V op kan leveren. Weliswaar kan deze spanning minimale stromen genereren, zodat deze niet schadelijk zal zijn voor mensen. Maar de spanning is hoog genoeg om vernielingen aan te kunnen richten in computers. Een computer werkt immers op een spanning van 5 V tot en met 15 V, dus zal een spanning van 4.000 V gegarandeerd een aantal onderdelen grondig vernietigen. Verwijdering van de statische electriciteit gebeurt door de kast van de computer aan te raken (let hierbij wel op de voedingsconnectoren, want het netsnoer zit nog in het stopcontact).
Vervolgens moet men ook al het gereedschap en de losse harde schijf even met de kast van de computer verbinden. Waar men nadien goed op moet letten is dat men het gereedschap en de harde schijf op een plek neerlegt waar geen statische electriciteit aanwezig is (het liefst op een anti-statische mat). Daarna kan men als laatste onderdeel het netsnoer verwijderen.

Stap 3
Deze stap is de voorbereiding voor het installeren van de harde schijf. Hier is het de bedoeling dat alle zaken, die in de weg zitten voor de plaatsing van de harde schijf in de computer, tijdelijk verwijderd worden. Hierbij kan men denken aan een simpel afdekplaatje van de computer of

de verwijdering van een floppy disk als de harde schijf hier de vervanging van is.
In figuur 3/19.5-12 is de plaats getekend van de houder, waarin de harde schijf aangebracht moet worden. Deze positie kan van computer-type tot computer-type afwijken, maar in ieder geval zal er steeds ergens in de kast een plaats gereserveerd zijn waarin de harde schijf kan aangebracht worden.

Stap 4
Deze stap is de voorbereiding voor het plaatsen van de controller kaart. Dit betekent dat voor die computers waar er een kaart in het meest rechtse slot zit, deze een slot zal moeten verschuiven (naar links dus). Dit is niet van toepassing als de kaart die aanwezig is een gecombineerde controller kaart is. Als men een tweede (of derde) controller kaart in de computer wenst te plaatsen, dan dient men deze ten allen tijde in een slot te plaatsen dat zover mogelijk aan de rechterzijde van de computer zit (vooraanzicht). De reden hiervoor is de vaak korte bekabeling die men ter beschikking heeft.

Stap 5
In deze stap wordt de controller kaart geïnstalleerd. Zaken die hier gedaan moeten worden zijn:
- kabel aansluitingen;
- jumpers instellen;
- plaatsing van de controller kaart in het slot.

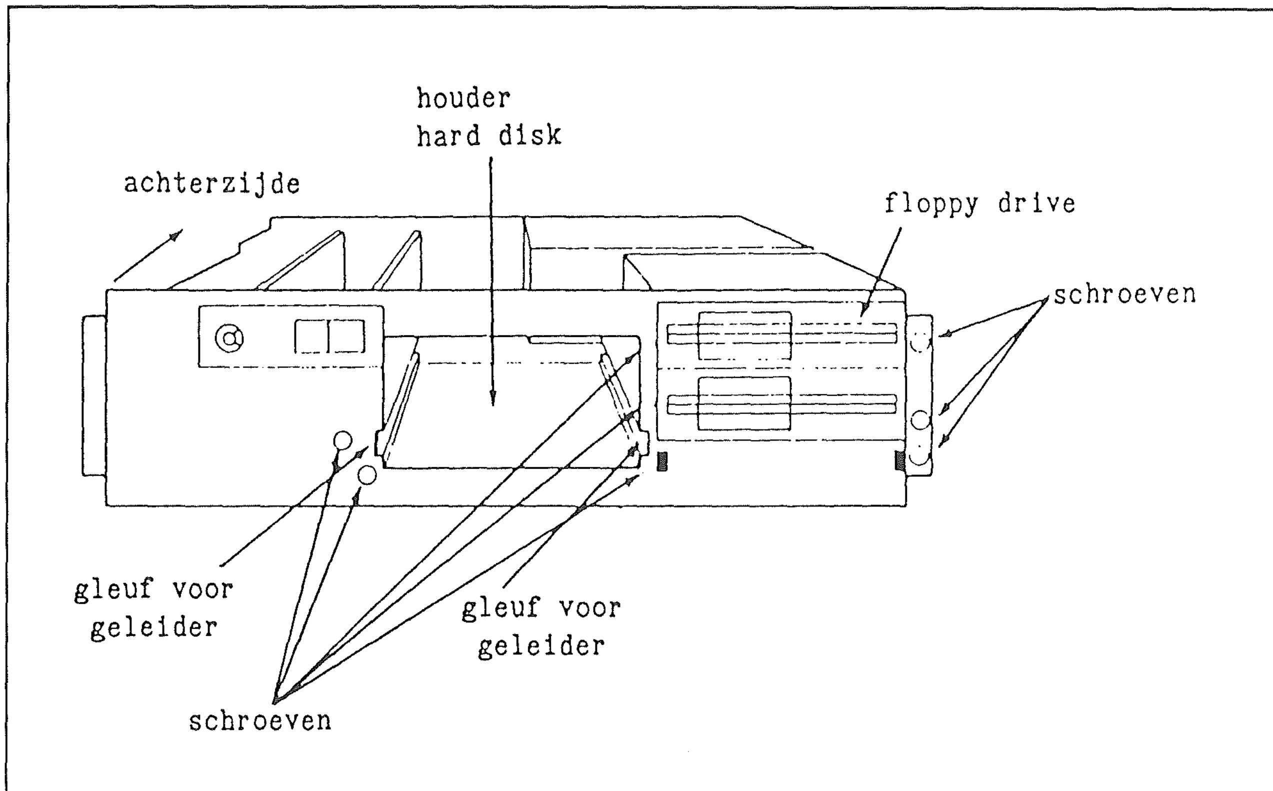
Op een controller kaart zitten een aantal connectoren, maar de connectoren waar het nu om gaat zijn de connectoren die bij de voorzijde van de computer terecht komen als de kaart geplaatst is. De kabels voor het aansluiten van de harde schijf zien eruit zoals geschetst in figuur 3/19.5-13. De meest rechtse connector komt op de controller kaart. Afhankelijk van de instellingen die bij aankoop op de controller kaart staan kan het nodig zijn om jumpers op de kaart te wijzigen. Als men dat wenst te doen, dan moet dat nu gebeuren. Nadien kan men de controller kaart in het daartoe voorziene slot plaatsen.

Stap 6
Dit is alleen nodig bij de oude PC en XT computer, want deze werken nog veelvuldig met jumpers. Om een correcte instelling te maken moet men het handboek van de computer ter beschikking hebben. Men zoekt de jumper instellingen op en zet vervolgens de jumpers in de goede positie.

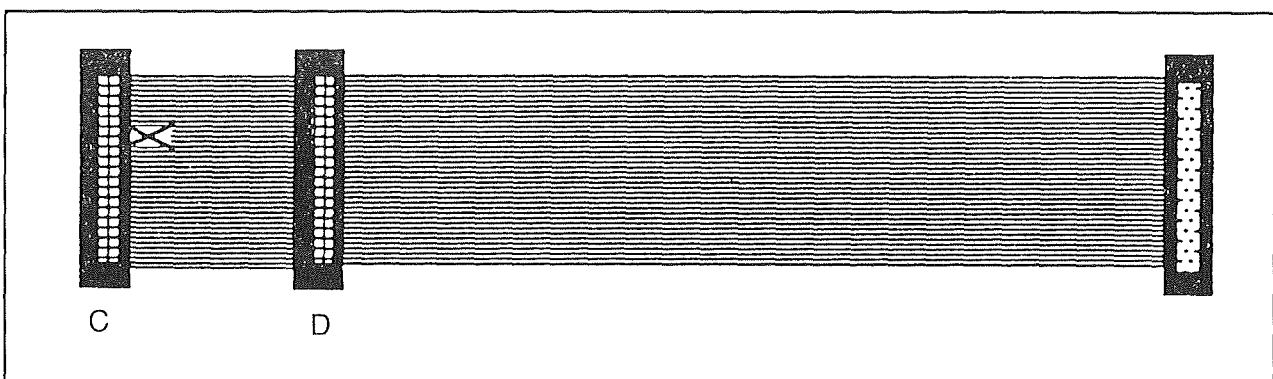
Stap 7
Deze laatste stap betreft de installatie van de harde schijf zelf. Items waar naar gekeken moet worden zijn:
- kabel aansluiting en terminator afsluiting;
- voeding aansluiting;
- aarde aansluiting;
- plaatsing van de harde schijf.

In figuur 3/19.5-13 staat de kabel van de harde schijf getekend. De meest linkse connector is voor de aansluiting op de eerste harde schijf ("Drive C"), terwijl de middelste connector bedoeld is voor de tweede harde schijf ("Drive D"). Een terminator is een afsluiting van de kabel. Deze wordt gebruikt om ervoor te zorgen dat er geen signaalvervorming (of echo) optreedt in de kabel. Op "Drive C" dient altijd een terminator te zitten en op "Drive D" moet deze verwijderd worden als "Drive C" aanwezig is.
Terminators zijn vaak weerstandsblokken die vlakbij de kabel aansluiting op de harde schijf zitten.

19.5 Principes van gegevensopslag op harde schijven



Figuur 3/19.5-12: De plaats van de houder voor de harde schijf in de behuizing van een gemiddelde PC.



Figuur 3/19.5-13: De kabel van de harde schijf controller.

Vervolgens dient men de voedingsplug op de harde schijf aan te sluiten. Daarna dient men een extra aarde verbinding op de harde schijf te plaatsen. Dit gebeurt vaak door een oogje van een kabeltje, die aan de binnenkant van de computer vastzit, aan de zijkant van de harde schijf te schroeven. Als alle kabels correct aangesloten en netjes weggewerkt zijn, kan men de harde schijf in de houder monteren.

Eindmontage

Als men al deze stappen doorlopen heeft, dan kan men de computer weer volledig aansluiten. Hier dient men de systeemkast nog niet vast te schroeven.
Daarna kan men de computer aanzetten om te kijken of alles functioneel is. Mocht dit niet het geval zijn, dan dient men de storing op te lossen. Als daarna alles goed functioneert dan kan men de systeemkast dichtschroeven en is de installatie voltooid.

De softwarematige preparatie van de harde schijf

Inleiding

Nadat de harde schijf in de systeemkast is gemonteerd, moet de nu nog volledig maagdelijke schijf worden voorbereid op het ontvangen van gegevens. Dat betekent het fysiek opdelen van de harde schijf in fysieke blokken die geschikt zijn voor het schrijven en lezen van data.
Dit onderdeel van de installatie valt uiteen in de volgende handelingen:

- drive type instellen;
- low level formatteren;
- partitionering;
- drive benaming toekennen.

Drive type instellen

Iedere drive heeft een typenummer die voor de computer aangeeft met welke harde schijf gewerkt wordt.
Dit typenummer definieert onder andere de capaciteit van de schijf, het aantal fysieke schijven dat aanwezig is en het aantal koppen.
Deze gegevens worden in het CMOS-geheugen van de computer opgeborgen, zodat het systeem bij het opstarten weet welke harde schijf aanwezig is.

Initialiseren

Het initialiseren van het drive type gebeurt op de volgende manier:

- PC of XT controller met AUTOCONFIGURE.ROM
  DOS commando DEBUG. Voor de meeste controllers geldt het commando "G=C800:0005".
  Voor de OMTI controller geldt "G=C800:0006" en voor de Paterson Labs Hdc-100/200 "G=C800:0008".
- XT controller
  Verander de jumper settings van de controller.
- AT controller
  Draai het SETUP programma.

Low level

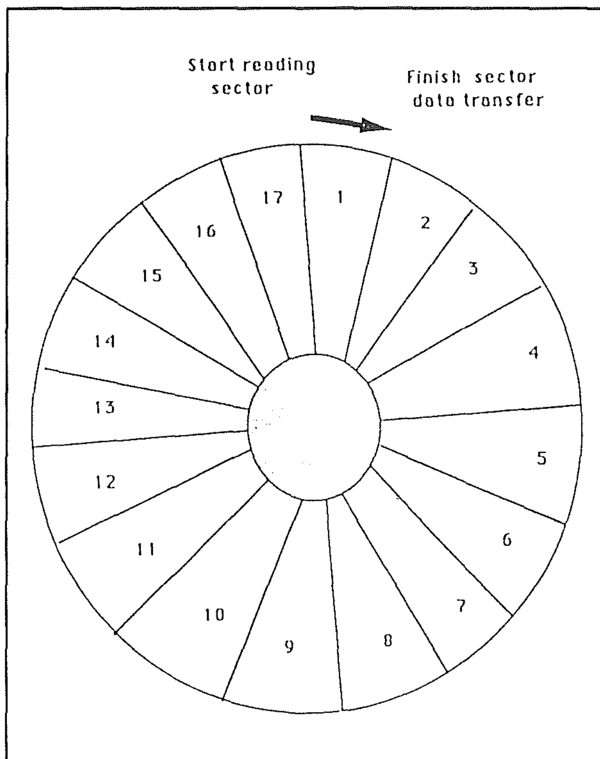
Voor het low level formatteren van de harde schijf geldt:

- XT controller met AUTOCONFIGURE.ROM
  Ga verder met het programma dat in de vorige stap opgestart is.
- Overige controllers
  Draai het IBM Advanced Diagnostics of HDDIAG programma.

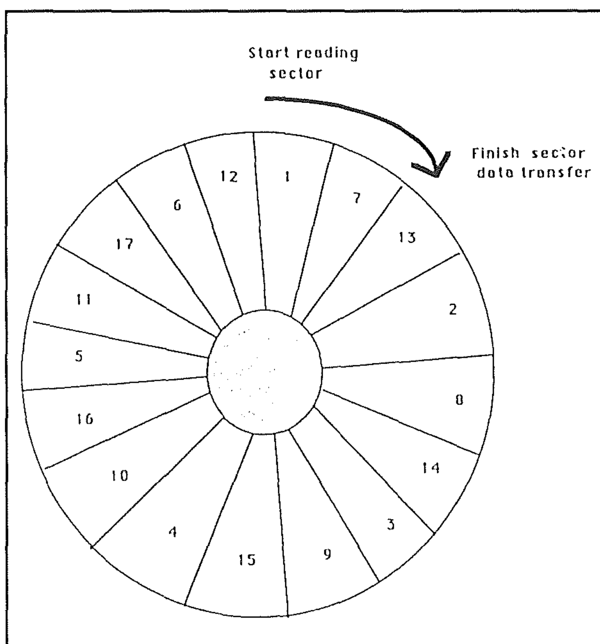
Interleave

Bij het low level formatteren van de harde schijf moet men de interleave van de harde schijf instellen. Een harde schijf is ingedeeld in sectoren. Deze taartsegmenten op het oppervlak van de schijven zijn de kleinste opbergplaatsen voor de gegevens (zie verder). Interleave wil zeggen het aantal stappen dat een registratiekop moet doen voordat de volgende sector gelezen kan worden.

19.5 Principes van gegevensopslag op harde schijven



Figuur 3/19.5-14: Het lees- en schrijfschema bij interleave 1:1.



Figuur 3/19.5-15: Het lees- en schrijfschema bij interleave 3:1.

De harde schijf draait snel rond, hetgeen betekent dat de registratiekop het begin van de volgende sector al voorbij is voordat deze kan beginnen met schrijven/lezen.

De meest gangbare interleave instellingen zijn:

- interleave 3:1 voor een PC of XT computer;
- interleave 2:1 voor een AT computer;
- interleave 1:1 voor een intelligente controller met een 16 of 32 bit datapad, zoals tegenwoordig standaard wordt toegepast bij Pentium-systemen.

In de figuren 3/19.5-14 en -15 zijn de sectorindelingen getekend voor respectievelijk interleaving 1:1 en 3:1. Bij 1:1 liggen de geschreven sectoren achter elkaar.

Bij 3:1 liggen er steeds twee sectoren tussen twee opeenvolgend geschreven of gelezen sectoren.

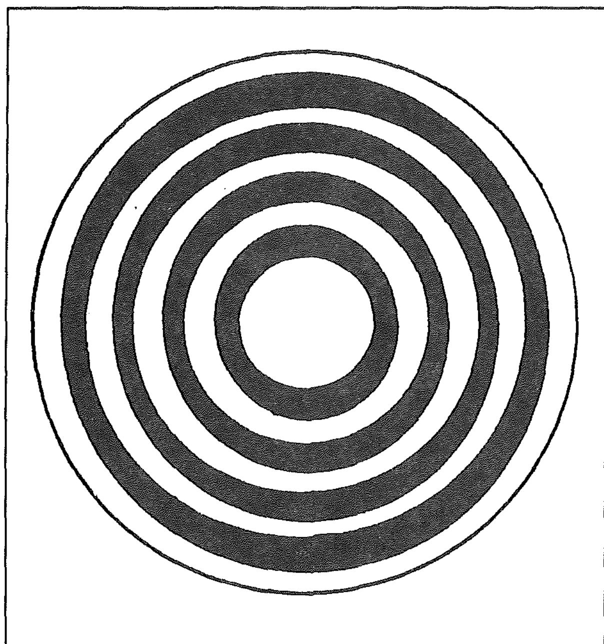
Partitionering

Partitionering van een harde schijf betekent het indelen van de schijf in verschillende logische stukken. Zo kan men bijvoorbeeld kiezen om een DOS-partitie, een Unix-partitie en een WIN95-partitie te maken.

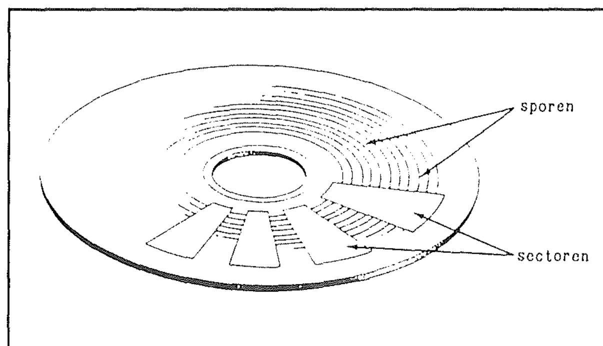
Iedere partitie wordt door het systeem als het ware herkend als afzonderlijke harde schijf. Zo kan men dus op een grote harde schijf drie partities maken, die dan bijvoorbeeld de DOS-benamingen C:, D: en E: kunnen krijgen. Er bestaan verschillende partitioneringsprogamma's, maar het meest bekende is het bij MS-DOS geleverde FDISK.

Als laatste stap moet men het MS-DOS commando FORMAT draaien om alle partities van de harde schijf te formatteren en om een drive letter aan iedere partitie tot te kennen.

19.5 Principes van gegevensopslag op harde schijven



Figuur 3/19.5-16: De spoorindeling van de harde schijf.



Figuur 3/19.5-17: De sectorindeling van de harde schijf.

Formatteren

De magnetische laag van de harde schijf is in de originele vorm onbruikbaar voor een computer. Om deze bruikbaar te maken moet een harde schijf geformatteerd worden. Het formatteren van een harde schijf betekent het indelen van de harde schijf in sporen (ook wel track's genoemd) en sectoren. Een spoor is een concentrische cirkel op de harde schijf, zie figuur 3/19.5-16. In deze figuur is de harde schijf getekend (de buitenste ring) met daarbinnen een aantal sporen (donkere cirkels). Dit is een theoretische voorstelling, want een harde schijf bestaat natuurlijk niet uit maar vier sporen, maar uit vele honderden sporen. Een sector is een onderdeel van een spoor, zie figuur 3/19.5-17. In deze figuur staat een aantal lege vlakken die ieder een sector voorstellen. Dit is echter plastisch omschreven, omdat de sectoren aaneengesloten op ieder spoor aanwezig zijn.

Als men een harde schijf indeelt in sectoren, dan is het de harde schijf controller die bepaalt hoeveel sectoren er op een spoor aanwezig zijn.

Er bestaan vier standaarden:

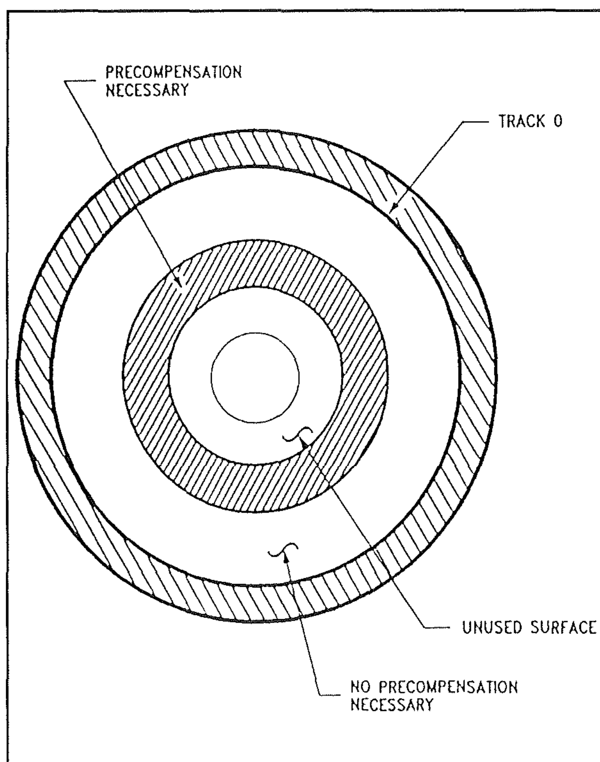
- ST412/ST506 met MFM, 17 sectoren per spoor;
- ST412/ST506 met RLL, 26 sectoren per spoor;
- SCSI, 32 tot 64 sectoren per spoor;
- ESDI, 32 tot 64 sectoren per spoor.

Park- en precompensation zône

Bij het formatteren van een harde schijf worden nog twee gebieden vastgelegd, namelijk de parkeerzône en de precompensation zône. Figuur 3/19.5-18 geeft deze verschillende zône's weer. De parkeerzône is een gebied waar de harde schijf zijn koppen kan parkeren bij stilstand. Dit is zeer nuttig als bijvoorbeeld de computer verplaatst wordt. Bij het vervoer kunnen de koppen gaan trillen, waardoor deze op de schijf zelf terecht komen en dit kan resulteren in een beschadiging van de magnetische laag. Het zal wél zonder nadere toelichting duidelijk zijn dat dit niet mag gebeuren op een gebied waar informatie staat! Er zijn twee parkeerzône's, namelijk aan de buitenrand (grof gearceerd in figuur 3/19.5-18) en aan de binnenrand (het stuk "unused surface").

19.5 Principes van gegevensopslag op harde schijven

De precompensation zône is een gebied waar dezelfde hoeveelheid data veel compacter op de betreffende sporen weg wordt geschreven. Dit is noodzakelijk omdat het aantal sectoren van een spoor niet verandert, maar de diameter van het spoor wél. De lengte van een sector neemt dus af naarmate men het middelpunt van de schijf nadert, maar toch moet dezelfde hoeveelheid informatie in deze kleinere lengte worden geschreven.



Figuur 3/19.5-18: De volledige zône-indeling van een harde schijf, met de parkeer- en precompensation zônes.

Moderne technologieën

Tegengestelde eisen

Aan moderne harde schijven worden tegengestelde eisen gesteld. Enerzijds moeten deze schijven steeds grotere capaciteiten krijgen, want een modern bedrijfssysteem plus een handjevol applicaties kost al snel een slordige 500 MB aan schijfruimte. Een harde schijf met een capaciteit van 2 GB zal binnen niet al te lange tijd standaard zijn in iedere gemiddelde computerconfiguratie. Anderzijds moeten de schijven echter steeds kleiner worden, omdat inbouw in steeds compactere draagbare apparatuur noodzakelijk is.
Grotere capaciteit op steeds minder oppervlak kan uitgedrukt worden in de gegevensdichtheid van de harde schijf. Een van de belangrijkste eigenschappen van een harde schijf is dan ook deze gegevensdichtheid.

Gegevensdichtheid

Bij de data-opslag op een harde schijf bestaat er een zeer complexe samenwerking tussen enerzijds het oppervlaktemateriaal van de schijf en anderzijds de constructie van de kop(pen). Met de standaard technologieën bereikt men op dit moment bij 3.1/2 inch harde schijven gegevensdichtheden van ongeveer 500 MB per vierkante inch. Hierbij is de afstand tussen naast elkaar liggende sporen gelijk aan 3 μm, terwijl de magnetische lengte van een bit slechts 0,5 μm bedraagt.

De koppen

De koppen vormen nog steeds het moeilijkste onderdeel van de harde schijf. Om zowel spoorafstand als bitlengte te kunnen minimaliseren, moeten de koppen zo klein mogelijk zijn en moet hun massa geminimaliseerd worden. Bovendien moet de afstand tussen de kop en de harde schijf zo klein mogelijk zijn, om zo min mogelijk magnetisch strooiveld op te wekken. Vroeger werden koppen van ferriet gemaakt, deze techniek werd echter vrij

snel verdrongen door de dunne-film technologie. Bij deze koppen wordt het spoeltje, waardoor de elektrische stroom vloeit die het magnetisch veld opwekt, gedrukt op een zeer dunne en flexibele kunststof drager. Beide systemen werken volgens het inductieve principe, waarbij een elektrische stroom die door het spoeltje wordt gestuurd een magnetische veld opwekt. Maar omgekeerd zal er in het spoeltje, dat door een magnetisch veld beweegt, een kleine spanning worden gegenereerd. Hoewel met de dunne-film techniek zeer kleine kop-constructies mogelijk zijn, zijn de eisen die moderne harde schijven stellen echter zo hoog, dat ook deze technologie niet meer voldoet.

Nieuwe technologieën

Om aan de moderne eisen, wat gegevensdichtheid betreft, tegemoet te komen, zijn in de voorbije jaren twee nieuwe interessante technologieën ontwikkeld:

- de MR-technologie heeft betrekking op de constructie van de koppen;
- de PRML-technologie heeft betrekking op de manier waarop het van de schijf gelezen signaal verwerkt wordt.

Beide technologieën maken gegevensdichtheden tot ver boven de 1 GB per vierkante inch mogelijk, terwijl de verwachting is dat rond het jaar 2000 een gegevensdichtheid van niet minder dan 10 GB per vierkante inch tot de mogelijkheden hoort.

De MR-technologie

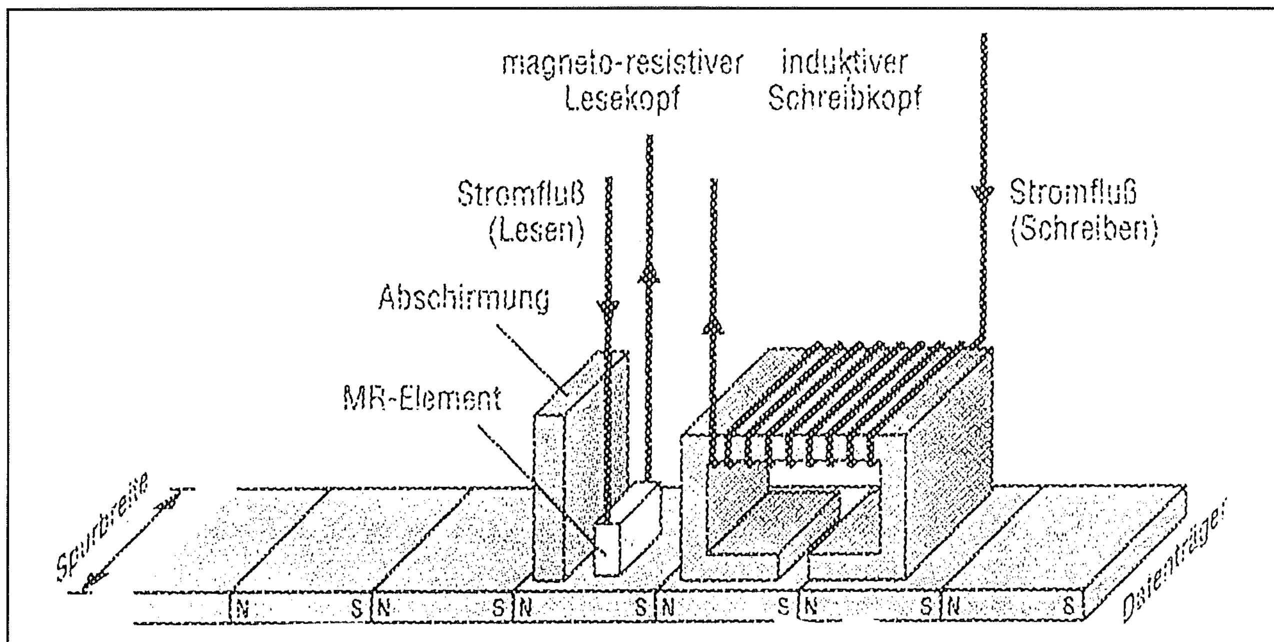
Het magneto-resistieve principe

De allermodernste koppen werken volgens het MR-principe. Deze afkorting is afgeleid van het woord "magneto-resitief". Sommige stoffen hebben de eigenschap dat zij hun soortelijke weerstand veranderen onder invloed van een magnetisch veld. Dit fysische verschijnsel staat ook bekend als "Hall-effect".
Weerstandsvariaties tot 2 % kunnen, zelfs bij de zeer zwakke magnetische velden die bij harde schijven aan de orde zijn, gemakkelijk gegenereerd en nadien gededecteerd worden.

De MR-kop

MR-koppen die volgens het beschreven principe werken, zijn in feite samengesteld uit twee verschillende systemen. Voor het schrijven van de signalen worden nog steeds inductieve koppen gebruikt, opgebouwd volgens de dunne-film technologie. Voor het lezen wordt echter een magneto-resitieve kop gebruikt. In figuur 3/19.5-19 is een schematische voorstelling van de samenstelling van een dergelijke kop getekend.
Het MR-element is slechts 30 nm dik, oftewel 0,000.000.030 meter! Dank zij deze kleine afmeting is het element veel beter in staat de magnetische informatie uit zeer smalle sporen af te lezen dan de traditionele inductieve leeskoppen.
Uit natuurkundige overwegingen moet het MR-element op hetzelfde elektrische potentiaal staan als het oppervlak van de schijf.
Is dat niet het geval, dan zouden er bij een "headcrash", het verschijnsel dat de kop te pletter slaat op de schijf, grote stromen gaan lopen tussen schijfoppervlak en kop, hetgeen tot vernietiging van het MR-element zou kunnen voeren. Om dit probleem op te lossen moet er in de MR-kop een afleidingscontact geïntegreerd zijn, dat er voor zorgt dat er nooit een potentiaalverschil tussen kop en plaat kan ontstaan.

19.5 Principes van gegevensopslag op harde schijven



Figuur 3/19.5-19: De samenstelling van een moderne MR-kop.

Voordelen van de MR-technologie

De MR-koppen hebben tal van voordelen, waarvan de voornaamsten zijn:

- het uitgelezen signaal is onafhankelijk van het toerental van de schijf;
- door de zeer kleine afmetingen is de spoorgeleiding minder kritisch;
- betere signaal-ruis verhouding;
- lezen-na-schrijven mogelijk.

Toerental onafhankelijk

Een groot voordeel van de MR-kop is dat het uitgelezen signaal volledig onafhankelijk is van het toerental van de schijf. Zelfs in stilstaande toestand zal de MR-kop het signaal, dat er onder aanwezig is, kunnen lezen. Dat betekent dat men in principe schijven met kleinere diameters kan toepassen. Een probleem bij rond draaiende schijven is immers dat de omtreksnelheid daalt naarmate men het middelpunt nadert. Omdat de traditionele inductieve leeskoppen een vrij hoge omtreksnelheid nodig hebben om een bruikbaar elektrisch signaal op te wekken, kan men niet te dicht bij het middelpunt van de schijf werken. Daar is immers de omtreksnelheid het laagst. Bij MR-koppen vervalt dit probleem geheel en al, zodat men veel kleinere schijfdiameters kan gaan gebruiken.

Minder kritische spoorgeleiding

De inductieve schijfkop heeft veel grotere afmetingen dan de MR-leeskop. Het gevolg hiervan is dat bij het lezen van de informatie de uiterst smalle leeskop altijd het volledige magnetische signaal zal oppikken. Het positioneringsmechanisme van de kop moet al erg grote afwijkingen vertonen, wil de MR-kop van het spoor lopen en minder signaal gaan afgeven.

Hogere signaal/ruis-verhouding

Een inductieve kop pikt niet alleen het magnetisch veld van de gegevens op de schijf op, maar ook allerlei stoorsignalen. Dit noemt men de ruis. Het zal duidelijk zijn dat deze ruissignalen tot leesfouten kunnen leiden. Ook op dit gebied hebben

19.5 Principes van gegevensopslag op harde schijven

MR-koppen grote voordelen. De signaal/ruis-verhouding is ongeveer twee tot vijf maal beter dan bij de traditionele inductieve koppen.

Lezen-na-schrijven
Doordat ieder kop afzonderlijke systemen heeft om te schrijven en te lezen, kan men heel eenvoudig een lees-na-schrijf principe gaan toepassen. Hierbij leest de MR-kop onmiddellijk na het schrijven de gegevens weer uit, zodat foutcorrectietechnieken zeer snel kunnen ingrijpen als er foutieve gegevens geschreven worden.

DSMR-technologie
Een inductieve kop werkt symmetrisch. Voor de opgewekte elektrische spanning maakt het niet uit of de kop door een zuid/noord dan wel door een noord/zuid magnetisch veld beweegt. Bij een MR-kop ligt dit echter anders. In beide gevallen zal de kop signalen van verschillende grootte genereren. Dit is lastig, omdat het dan problematisch wordt om een goed spoorgeleidingssysteem te ontwerpen.
Vandaar dat sommige fabrikanten zijn overgegaan op de zogenoemde DSMR-techniek. Dit letterwoord is de afkorting van "Dual Stripe Magneto Resistive". Hierbij worden in iedere kop twee MR-elementen ondergebracht, die in anti-serie worden geschakeld. Het gevolg hiervan is dat het ene element een groot signaal zal genereren bij een noord/zuid magnetisch veld en de tweede kop een klein signaal. Maar, anderzijds, de eerste kop zal een klein signaal genereren bij een zuid/noord veld en de tweede kop een groot signaal. Door de anti-serie schakeling zal de combinatie van beide MR-elementen in beide gevallen alleen een verschilsignaal genereren. Deze verschilsignalen zijn even groot.

De DSMR-technologie heeft als bijkomstig voordeel, dat de signaal/ruis-verhouding nog groter wordt en dat de noodzaak van een afleidingscontact vervalt. Men kan nu immers het middelpunt van de serieschakeling van beide MR-elementen met de massa verbinden, waardoor nooit ofte nimmer grote spanningsverschillen tussen de kopconstructie en het oppervlak van de schijf kunnen ontstaan.

De GMT-technologie
Een door IBM ontwikkelde technologie, die GMT wordt genoemd, verbetert de specificaties van MR-koppen in belangrijke mate. GMT staat voor "Giant Magneto Resistive". Hierbij wordt de MR-kop niet samengesteld uit een welbepaald metaal, maar uit een sandwich. De meest gebruikte constructie bestaat uit afwisselende lagen van een magnetische nikkel/ijzer-legering en niet-magnetisch zilver. Deze lagen worden opgedampt op een substraat (drager) van tantalium. Nadat alle lagen zijn opgedampt, wordt het substraat verhit tot ongeveer 300 °C. Hierbij dringen de zilver atomen in de nikkel/ijzer legering binnen. Hierdoor ontstaan zeer kleine zogenoemde magnetische domeinen in het materiaal. Het gevolg is dat het materiaal veel gevoeliger wordt voor magnetische velden, waardoor weerstandsvariaties van ongeveer 6 % in de praktijk haalbaar zijn. Dat is dus ongeveer drie keer meer dan bij de traditionele MR-koppen. Het gevolg is dat de spoorbreedte op het oppervlak van de harde schijf kleiner kan worden en dat de gegevensdichtheid toeneemt.
Bij een in 1994 ontwikkelde experimentele harde schijf met GMR-koppen kon de spoorafstand tot 0,5 μm verkleind worden.

19.5 Principes van gegevensopslag op harde schijven

In diverse laboratoria wordt gewerkt aan sandwich-constructies van andere metalen, in de hoop dat de gevoeligheid van het aldus ontstane GMR-materiaal nog verder opgevoerd kan worden. De verwachtingen zijn dat reeds in het jaar 2000 harde schijven verkrijgbaar zijn met koppen volgens de GMR-technologie, met gegevensdichtheden van rond 10 GB per vierkante inch.

De praktijk

IBM heeft reeds in 1989 een eerste experimentele MR-schijf ontworpen. Hiermee konden gegevensdichtheden van niet minder dan 1 GB per vierkante inch bereikt worden. In 1991 kwam de eerste commerciële uitvoering van een harde schijf met MR-koppen in de handel. Diverse fabrikanten zijn op dit moment volledig omgeschakeld naar de MR-technologie voor al hun schijven met een capaciteit die groter is dan 1 GB.

De PRML-technologie

Inleiding

Hoe kleiner men de breedte van de magnetische sporen op de harde schijf maakt, hoe meer sporen men op een bepaalde schijfdiameter kan aanbrengen en hoe groter de capaciteit van de schijf wordt. Er zijn echter grenzen gesteld aan de minimale breedte van een spoor. Hoe smaller het spoor wordt, hoe kleiner het signaal zal zijn dat door de leeskoppen wordt opgepikt. Op een bepaald moment wordt het signaal zo klein dat het verloren dreigt te gaan in de ruis, die in ieder elektronisch systeem aanwezig is. De signaal/ruisverhouding van de kop wordt dan zo slecht dat de achtergeschakelde elektronica er met geen mogelijkheid meer een mooi digitaal signaal van kan maken.

Digitale signaal processing

De moderne digitale technieken staan echter voor niets! Men heeft in de voorbije vijftien jaar immense vorderingen gemaakt met het ontwikkelen van elektronische systemen, die in staat zijn zeer kleine signalen terug te winnen uit ruis die veel groter is dan de signalen zélf. Dergelijke technieken worden bijvoorbeeld al jaren gebruikt om de beelden, die ruimteverkenners die op vele tientallen miljoenen kilometer afstand foto's van het oppervlak van planeten maken, te bewerken tot herkenbare plaatjes. Dergelijke technieken berusten uiteraard op een heleboel fundamentele wiskunde.

Algoritmen

De wiskundige technieken die men gebruikt bij dergelijke signaalherkenningstechnieken berusten op twee wiskundige algoritmen, die al tamelijk bejaard zijn.
In 1963 werd een algoritme ontwikkeld, dat bekend staat onder de naam "Partial Response Signalling". Dit algoritme probeert niet individuele bits uit ruis te bevrijden, maar kijkt naar de opeenvolging van een aantal bits, waarbij gepoogd wordt bepaalde bit-combinaties te herkennen. Iets dergelijks gebeurt er ook in de menselijke hersenen. Deze zijn immers in staat een zeer onleesbaar geschreven woord toch te identificeren aan de hand van een paar wél leesbare letters.
In 1973 werd het "Maximum Likelihood Detection"-algoritme ontwikkeld. Dit gaat er van uit dat van woorden, die niet als dusdanig herkend kunnen worden, toch de betekenis te achterhalen is door te kijken naar de woorden die er voor staan en die er na komen.

Het PRML-algoritme

Uit beide genoemde algoritmen werd in 1990 door IBM een nieuw algoritme samengesteld, dat door het leven gaat onder de naam "Partial Response Maximum Likelihood", afgekort tot PRML. De elektronische vertaling hiervan is zeer complex, maar kan dank zij de moderne µm-technologieën toch tot een handzame hoeveelheid onderdelen terug gebracht worden. Er werd een speciale signaalprocessor voor dit algoritme ontwikkeld, die in staat is ongeveer 60 MB/s aan gegevens te verwerken. De processor is in staat 6 bit brede vensters te analyseren en hieruit iedere 10 ns een bit te berekenen. Dank zij de PRML-techniek is men in staat de breedte van de sporen die op het oppervlak van een harde schijf geschreven worden aanzienlijk te reduceren, zonder dat de gegevens foutief uit de ruis gelezen worden.

Omcodering

Een nadeel van de PRML-technologie is dat de van de computer afkomstige gegevens omgecodeerd moeten worden. Hierbij zijn omcoderingen van 8/9 en 16/17 ontwikkeld. Dat wil zeggen dat voor het bewaren van 8 data-bits uit de computer er 9 bits op de harde schijf geschreven moeten worden (of natuurlijk 17 voor 16 computer-bits). Desondanks biedt de PRML-technologie grote voordelen, waardoor steeds meer fabrikanten deze gaan toepassen.

De praktijk

Op dit moment leveren diverse fabrikanten harde schijven, die de gelezen gegevens bewerken met het PRML-algoritme. Een paar voorbeelden:

- de Allegro-2 van Fujitsu met een capaciteit van 4,3 GB;
- de SuperStore 2000LP van Hewlett Packard met een capaciteit van 2,14 GB;
- de Ultrastar XP van IBM met een capaciteit van 4 GB;
- De Atlas II van Quantum met een capaciteit van 9,1 GB.

3/19.7

Principes van de toetsenbord elektronica

Inleiding

Een gewoon randapparaat

Voor de PC is het toetsenbord niet meer of minder dan een van de vele perifere schakelingen, die op het systeem kan worden aangesloten. De processor uit de PC beschouwt het toetsenbord als een seriële perifere schakeling, die de gegevens over welke toets of toetsen worden ingedrukt onder seriële vorm aan de elektronica van de computer aanbiedt. Het toetsenbord bevat dus een elektronische schakeling die niet alleen detecteert welke toets op een bepaald moment wordt ingedrukt, maar deze detectie omzet in een seriële code die via de kabel waarmee het toetsenbord met de computer verbonden is wordt verzonden.

Evolutie

In de snel evoluerende wereld van de PC-elektronica is het toetsenbord een vrij stabiele factor.

In de loop der tijden zijn er in feite slechts drie verschillende toetsenborden ontwikkeld:

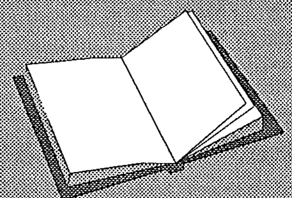
- het PC/XT-toetsenbord;
- het AT-toetsenbord;
- het MF II toetsenbord.

Hoewel er duidelijke verschillen tussen deze drie toetsenborden bestaan, is de basiswerking in grote lijnen hetzelfde gebleven.

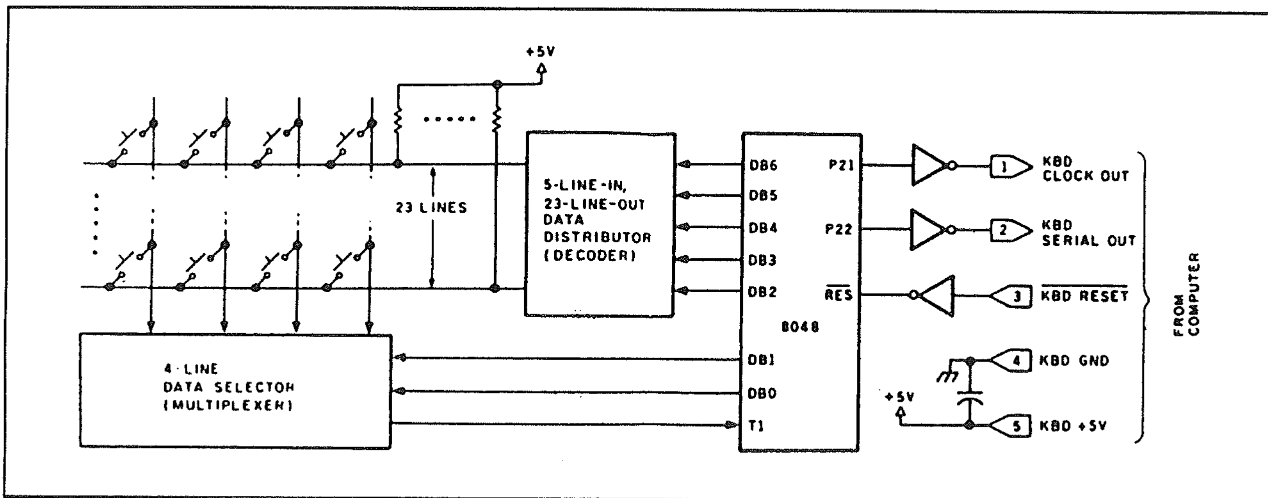
Het originele PC/XT-toetsenbord

De fundamentele werking van een toetsenbord kan het beste worden beschreven aan de hand van de oudste en eenvoudigste uitvoering, het PC/XT-toetsenbord. Dit werd door IBM ontwikkeld en werd standaard bij de origineel IBM-PC geleverd. Het schema van de elektronica van het originele IBM-toetsenbord is getekend in figuur 3/19.7-1. Hart van deze schakeling is een speciale microprocessor van Intel, de 8048. Dit IC is de keyboard-controller, de schakeling die het gehele toetsenbord bestuurt. Drie aansluitingen van dit IC zijn via inverterende buffers aangesloten op de KEYBOARD CLOCK, de KEYBOARD SERIAL OUT en de KEYBOARD RESET van de computer. Daarnaast moet de schakeling van het toetsenbord uiteraard gevoed worden, dus worden de +5 V systeemvoeding en de massa via de kabel naar het toetsenbord gevoerd.

LEES OOK:
Hoofdstuk 3/19.1
Hoofdstuk 3/19.8

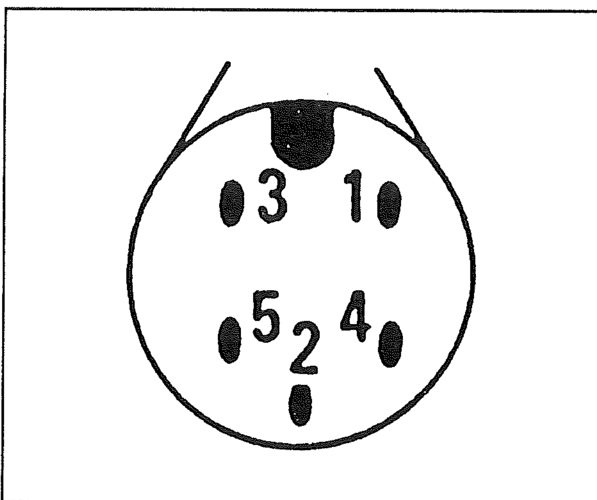


19.8 Principes van de toetsenbord elektronica



Figuur 3/19.7-1: De elektronica van het originele IBM-toetsenbord.

Deze vijf verbindingen worden via een gestandaardiseerde steker, een uit de audio-elektronica bekende vijfpolige DIN-steker, op een connector van de computer aangesloten. De aansluitings-code van deze steker is getekend in figuur 3/19.7-2.



Figuur 3/19.7-2: De aansluitingen van de gestandaardiseerde vijfpolige connector van het toetsenbord.

De pen-nummers corresponderen met onderstaande signalen:

- pen 1: KEYBOARD CLOCK;
- pen 2: KEYBOARD SERIAL OUT;
- pen 3: KEYBOARD RESET;
- pen 4: MASSA (GROUND);
- pen 5: +5 V.

De CLOCK-uitgang wordt gebruikt voor het synchroniseren van de seriële gegevens op de SERIAL OUT lijn. De RESET wordt gestuurd uit de start-routine van de PC-processor, hoewel dit zuiver elektronisch bekeken niet per se noodzakelijk is.

Toetsen-matrix

De toetsen zijn opgenomen in een zogenaamde matrixschakeling, opgebouwd uit 4 vertikale en 23 horizontale lijnen. Op ieder snijpunt van twee lijnen is het contact van een toetsschakelaar in de matrix opgenomen. Als men op deze toets drukt zal de horizontale met de vertikale lijn verbonden worden. De 23 horizontale lijnen zijn via weerstanden verbonden met de +5 V systeemvoeding en staan dus normaal op logisch "H". Daarnaast zijn deze lijnen aangesloten op de uitgangen van een multiplexer, die op zijn beurt wordt gestuurd uit vijf uitgangen van de 8048. Door op deze vijf uitgangen een bepaalde binaire code te zetten zal een van de horizontale lijnen door de multiplexer naar de massa getrokken worden en logisch "L"

19.7 Principes van de toetsenbord elektronica

worden. Op dat moment zijn de overige horizontale lijnen nog “H”. Als men op een toets drukt die op een snijpunt van deze horizontale lijn is opgenomen, dan zal de vertikale lijn die mede dit snijpunt vormt door het gesloten contact van de ingedrukte toets ook naar massa getrokken worden. Deze informatie wordt door een data-selector, gestuurd door de 8048, omgezet in een datapuls voor de 8048.
Uiteraard worden alle horizontale lijnen opeenvolgend door de multiplexer met de massa verbonden. Bij ieder stap in deze zogenaamde scan-procedure onderzoekt de 8048 of, en zoja welke, vertikale lijn laag is. Als een laag signaal wordt gedetecteerd wacht de 8048 enige honderdsten van een seconde alvorens verdere actie te ondernemen. Mechanische schakelaars vertonen namelijk de eigenschap dat de verende metalen lipjes die het contact vormen een bepaalde tijd gaan trillen, waardoor de elektrische verbinding vele malen achter elkaar wordt gesloten en weer geopend. Dit noemt men het bouncen van de schakelaar. Zolang deze mechanische trillingen optreden zal de spanning op de vertikale lijn niet stabiel zijn en zouden er foutieve conclusies kunnen worden getrokken. Na deze zogenaamde debounce-tijd berekent de 8048 uit de momentele code van de signalen die een horizontale lijn selecteren en het signaal dat wordt ontvangen op de vertikale lijnen welke toets wordt ingedrukt.
Deze informatie wordt in een buffer in de 8048 opgeslagen en wordt gebruikt om de zogenaamde scan-code onder seriële vorm naar de computer te sturen.

Autorepeat
De 8048 detecteert niet alleen wanneer een toets wordt ingedrukt, maar ook wanneer deze toets weer wordt losgelaten. Op dat moment wordt een andere scan-code naar de computer gestuurd. Daarnaast zal de schakeling de scan-code van een toets die langer dan een halve seconde ingedrukt blijft, repeterend naar de computer zenden.
Dit noemt men de auto-repeat functie van het toetsenbord.

Overige functies
De processor van het toetsenbord beschikt over enige primitieve fout-detectieroutines en is voorzien van een klein buffer-geheugen waarin 20 opeenvolgende toetsenbordaanslagen kunnen worden opgeslagen. Deze buffer wordt ingeschakeld op de zeldzame momenten waarop de hoofdprocessor van de PC niet in staat is een toetsenbordcode onmiddellijk te accepteren.

Opmerking
De beschreven schakeling geldt alleen voor de originele IBM-PC en de hiervan afgeleide volledig compatibele klonen. Er zijn echter tal van klonen op de markt gekomen, die een afwijkende toetsenbord schakeling hebben.

Het AT-toetsenbord
Het AT-toetsenbord werd gekenmerkt doordat er rechts een afzonderlijk toetsenbordje kwam voor de numerieke toetsen, met een vormgeving die doet denken aan het toetsenbordje van een eenvoudig rekenmachine. Ook werd een andere schakeling toegepast voor het decoderen van het toetsenbord. Bij deze reeks PC’s wordt een keyboard-encoder toegepast van het type 8042, 8741 of 8742. Deze controllers werken echter in grote lijnen op dezelfde manier als de 8046. Het grote verschil is echter dat er een beperkte mate van intelligentie aanwezig is. Deze con-

trollers kunnen worden geprogrammeerd, bijvoorbeeld om het toetsenbord volledig uit te schakelen of om de "Caps Lock" in te schakelen. Er is dus sprake van een beperkte mate van bi-directionele gegevensoverdracht tussen de elektronica van het toetsenbord en de elektronica van de PC.

Het MF II toetsenbord

Het AT-toetsenbord werd opgevolgd door het zogenoemde MF II toetsenbord. In feite is dit toetsenbord nu nog steeds de standaard van alle PC's. Dit toetsenbord werd uitgebreid met de functietoetsen F11 en F12 (die helaas nauwelijks door software worden gebruikt) en bovendien werden de besturingstoetsen opgenomen in een extra toetsenblokje. De 12 functietoetsen werden boven aan het toetsenbord op een rijtje geplaatst. Bovendien werden de vier cursortoetsen ook nog eens gegroepeerd en werden er dubbele "Alt"-, "Ctrl"- en "Shift"-toetsen opgenomen. Ook de toetsen "Print Screen" en "Pause" zijn kenmerkend voor een MF II toetsenbord.
Ook de drie LED's, waarmee de status van "Caps Lock", "Num Lock" en "Scroll Lock" wordt aangegeven, werden ingevoerd bij de introductie van het MF II toetsenbord. Elektronisch bekeken zijn er weinig verschillen tussen dit toetsenbord en zijn voorganger AT.

De scan-codes

Inleiding

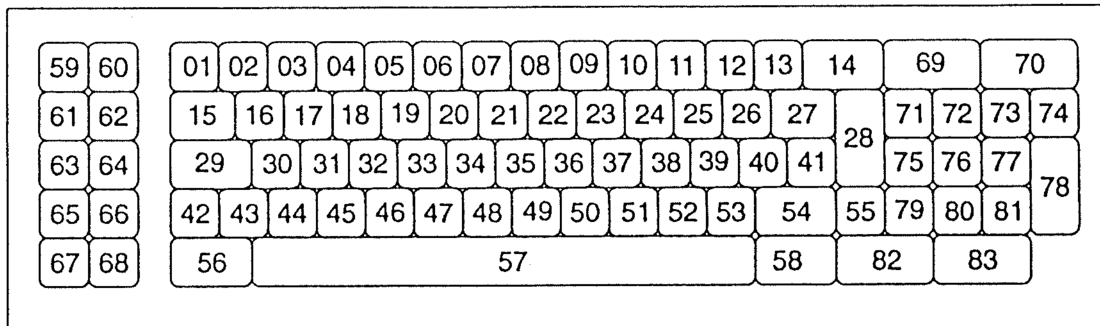
Iedere keer als de gebruiker/ster een toets op zijn of haar toetsenbord indrukt of loslaat, genereert de elektronica een code die uniek is voor iedere toets. Deze code wordt de scan-code genoemd. Opgemerkt moet worden dat deze scan-codes niets te maken hebben met de bekende en internationaal gestandaardiseerde ASCII-code die aan iedere letter, cijfer en leesteken een unieke 8 bit brede code toekent. Het verband tussen de scan-code en de ASCII-code wordt softwarematig vastgelegd door het bedrijfssysteem van de computer.
Dit heeft als groot voordeel dat het in principe mogelijk is het volledige toetsenbord te herdefiniëren zonder dat daarvoor de hardware van de computer gewijzigd moet worden. Op kleine schaal wordt van deze eigenschap gebruik gemaakt als men het toetsenbord van de computer aanpast aan de landstaal door het laden van een van de KEYB.EXE-files die standaard bij iedere versie van MS-DOS geleverd worden.

De scan-codes

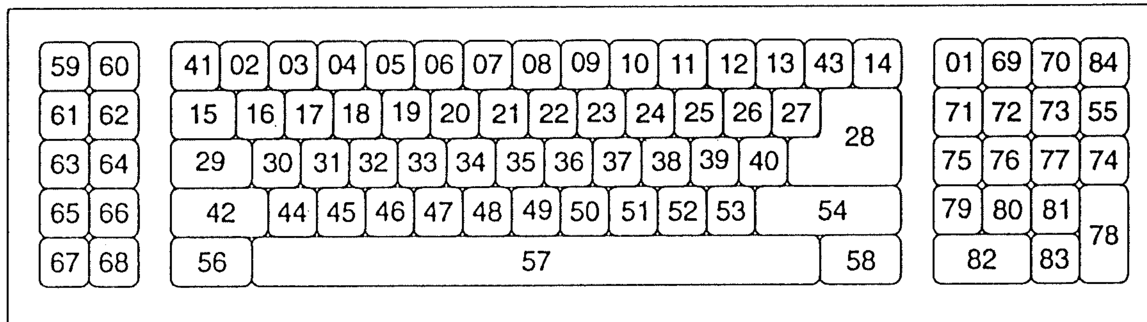
Alle toetsen wekken bij het indrukken een scan-code op die ligt tussen decimaal 1 en decimaal 83 (althans bij het PC/X-toetsenbord). Dit noemt men de make-code. Bij het loslaten van dezelfde toets zendt het toetsenbord een code uit die decimaal 128 hoger ligt. Het ligt voor de hand dat men hiervoor het woord break-code bedacht heeft. De vaste relatie tussen de make- en de break-code van een en dezelfde toets heeft een erg voor de hand liggende reden. Bij het drukken van een toets is bit 7 van de code steeds "L", bij het loslaten van de toets maakt de elektronica bit 7 "H". En zoals men weet heeft het zevende bit van een byte een decimale waarde van 128. Figuur 3/19.7-3 geeft een overzicht van de make-waarde van de scan-codes voor alle toetsen van de drie standaard toetsenborden PC/XT, AT en MF II.

19.7 Principes van de toetsenbord elektronica

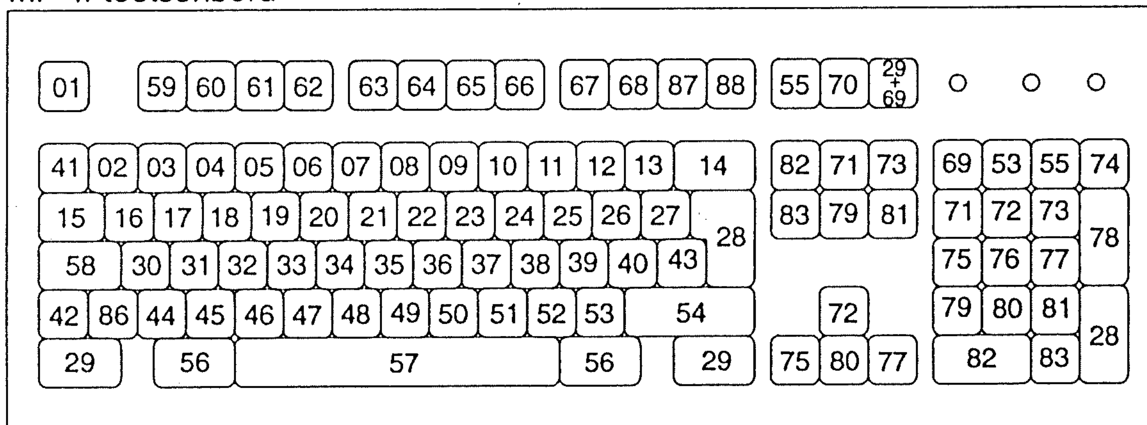
PC/XT-toetsenbord



AT-toetsenbord



MF II-toetsenbord



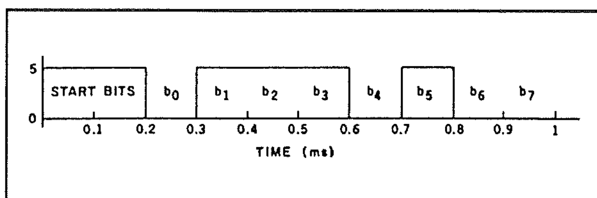
Figuur 3/19.7-3: De make-codes voor de drie toetsenborden.

Compatibiliteitsproblemen
De overgang van het PC/XT-toetsenbord naar het AT-toetsenbord kon probleemloos verlopen. Het AT-toetsenbord heeft immers maar één extra toets, namelijk "Sys Req". Hieraan werd een nieuwe scancode toegekend, namelijk 84. Voor de rest kon alles bij het oude blijven.
Bij de introductie van het MF II toetsenbord deden er zich wel problemen voor.

Dit toetsenbord beschikt immers over een heleboel nieuwe toetsen en bovendien werden een aantal toetsen dubbel uitgevoerd. Indien deze toetsen dezelfde scan-codes zouden krijgen, zou het systeem niet kunnen ontdekken of er op de linker dan wel de rechter "Alt"-toets wordt gedrukt. Het introduceren van een geheel nieuwe scan-code zou onoverkomelijke compatibiliteitsproblemen tot gevolg hebben. Daarom werden de scan-codes van de nieuwe toetsen uitgebreid. Er wordt een zogenoemd voorzet-byte aan toegevoegd, met als waarde $E0_{hex}$ of $E1_{hex}$. Daarna volgt de bekende make- of break-code. Het voorzet-byte $E1_{hex}$ wordt alleen verzonden bij het indrukken van de "Pause"-toets. Voor alle andere nieuwe toetsen wordt het byte $E0_{hex}$ verzonden.

Het verzenden van de scan-code

Zoals reeds gezegd, worden de bit's van de scan-codes onder seriële vorm over de KEYBOARD SERIAL OUT-lijn verzonden. Het protocol van deze seriële transmissie is getekend in figuur 3/19.7-4.



Figuur 3/19.7-4: Het protocol van de transmissie van één scan-code door het toetsenbord.

In rust, dus als er niet op een toets wordt gedrukt, is de spanning op deze lijn "L". Bij het begin van de scan-code transmissie wordt de lijn gedurende ongeveer 0,2 ms "H" gemaakt. Deze puls komt overeen met twee start-bits. Nadien worden alle 8 bits van het scan-byte in een ritme van 0,1 ms per bit op de lijn gezet, waarbij het laagste of LSB-bit het eerst aan de beurt komt. Een hoog bit wordt vertaald naar een spanningspuls van +5 V, een laag bit naar een spanning van 0 V. Let er dus op dat de bit-volgorde anders is dan deze die wordt gebruikt bij het schrijven van de inhoud van een byte op papier! Men schrijft immers van MSB naar LSB, de elektronica verzendt van LSB naar MSB! De code die in figuur 3/19.7-4 als voorbeeld wordt aangehaald is samengesteld uit "L-L-H-L-H-H-H-L", hetgeen staat voor decimaal 46, de scan-code voor de "C"-toets.

De elektronica in de computer

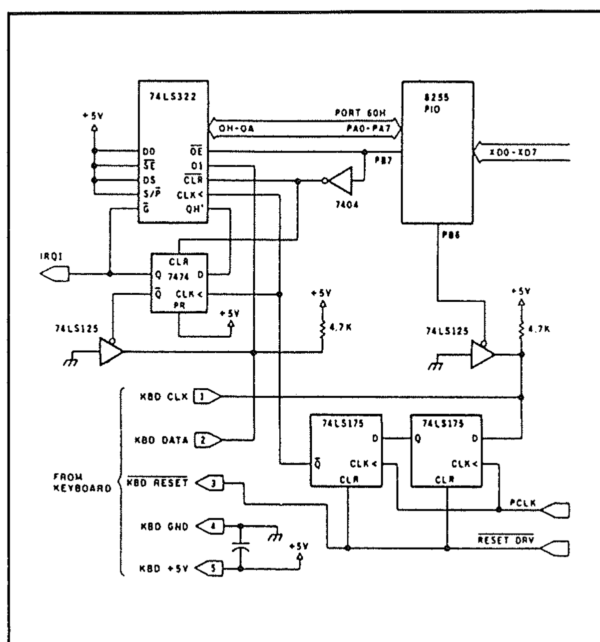
Inleiding

In principe wijken de seriële data die door het toetsenbord worden verzonden niet af van de standaard-manier waarop gegevens onder seriële vorm worden verstuurd. Men zou dus ook de standaard schakelingen kunnen toepassen voor het omzetten van deze seriële data in parallelle vorm. Zo is het heel goed mogelijk een toetsenbord-decoder te maken met een UART, hoewel men dan wel de polariteit van enige logische signalen zou moeten inverteren. UART's werken immers met een rustsignaal dat "H" is en niet "L" zoals dat van het toetsenbord.

De IBM-oplossing

IBM heeft echter in de oorspronkelijke PC voor een tamelijk eenvoudige logische schakeling gekozen, waarvan het blokschema getekend is in figuur 3/19.7-5.

19.7 Principes van de toetsenbord elektronica



Figuur 3/19.7-5: Het omzetten van de seriële gegevens van het toetsenbord in parallelle data en het aanvragen van een hardware-interrupt.

Het signaal op de KEYBOARD CLOCK-lijn wordt aangeboden aan twee in cascade geschakelde type-D flip-flop's in een 74LS175. Deze schakeling heeft tot taak de clock-pulsen over twee perioden te vertragen en het geheel te synchroniseren met de systeem-clock van de computer. Dat laatste signaal wordt immers aan de CLK-ingangen van de flip-flop's aangeboden en de informatie op de D-ingangen zal eerst bij de eerstvolgende positieve flank van de systeem-clock worden verwerkt. De seriële data worden aangeboden aan de seriële ingang van een 74LS322, een schuifregister met seriële ingang en parallelle uitgangen. De gegevens worden door de uit de flip-flop's afgeleide pulsen in het schuifregister geschoven. Als alle 8 bits in het register zitten wordt de carry-uitgang van deze schakeling "H" (vandaar de hoge start-bits bij het protocol!) en deze vraagt een interrupt aan bij de systeemprocessor via de IRQ1-lijn. De processor schakelt de interrupt-routine INT 9 in, die de parallelle gegevens uit het register inleest via poort A van de 8255 PIO. Via de PB7-lijn van deze schakeling wordt de 7474 flop-flop gereset en valt de interrupt request weg.

De verwerking van de scan-codes

De functie van het BIOS

De INT 9 routine leest de scan-code uit poort 96 en zet deze code om in een uit twee bytes opgebouwd woord. Het eerste byte wordt gevuld met de ASCII-code van de ingedrukte toets, het hoge byte bevat de scan-code van het toetsenbord. Bij speciale toetsen, zoals de functie-toetsen en de cijfertoetsen uit het numeriek bordje rechts op het toetsenbord, wordt het eerste byte echter leeg gelaten.
De op deze manier omgevormde toetsencodes worden nadien in een registertje in het geheugen geplaatst, de zogenaamde toetsenbord-buffer, die ligt tussen adres 0000_{hex} en adres $041E_{hex}$.

Shift, Alt, Ctrl en Lock

Tot nu toe is de noodzaak van twee scancodes per toets niet erg voor de hand liggend. Waarom moet de processor zowel op de hoogte gebracht worden van het indrukken van een toets (make-code) als van het weer loslaten daarvan (breakcode)? Vergeet echter niet dat het toetsenbord van de PC een ingewikkeld instrument is, waarbij de meeste toetsen veel meer dan één betekenis hebben. Het maakt heel wat uit of men bijvoorbeeld

19.8 Principes van de toetsenbord elektronica

alleen de "C"-toets indrukt of men dit doet terwijl de "Ctrl"-toets al ingedrukt is.
De processor moet in staat zijn zo'n dubbele toetsdrukken te detecteren en vandaar dat het zo belangrijk is dat er zowel make- als break-codes bestaan. Detecteert de processor immers een opeenvolging van de make-code van de "Ctrl"-toets, de make-code van de "C"-toets en de breakcode van de "Ctrl"-toets, dan weet het systeem dat beide toetsen gelijktijdig zijn ingedrukt en dat er iets anders wordt verwacht dan het schrijven van een letter C naar het video-geheugen.
Op een soortgelijke manier is de processor in staat te detecteren of de "Caps Lock"- of "Num Lock"-toetsen geactiveerd zijn. Treft de scan-routine een make-code voor een van deze toetsen aan, dan zal één bit van de adressen 417_{hex} en 418_{hex} gezet worden. Bij het interpreteren van een scan-code van een willekeurige toets onderzoekt de processor de status van deze bits. Aan de hand van de gevonden waarde weet de schakeling welke ASCII-code er in het twee bit lange reeds vermelde woord moet worden geschreven.
Bij het ontvangen van de break-code van de "Lock"-toetsen wordt het betreffende bit in de genoemde adressen weer gereset.

Speciale combinatie-toetsen
De make- en break-codes zijn ook noodzakelijk voor het interpreteren van speciale toets-combinaties die een directe actie van de computer verwachten. De meeste van deze acties worden door MS-DOS ondersteund, hetgeen niet wil zeggen dat een programmeur deze acties niet kan ontwijken.
Een overzicht van deze combinaties:

- "Ctrl" + "Break":
  De processor voert een INT 27 uit, waardoor het lopende programma wordt onderbroken. Het is uiteraard voor een programmeur mogelijk een eigen routine te schrijven die de INT 27 onderschept en de computer opdraagt iets anders te doen.
- "Ctrl" + "Num Lock":
  Het programma wordt onderbroken totdat er op een willekeurig andere toets wordt gedrukt.
- "Shift" + "Print Sceen":
  De processor schakelt een BIOS-routine in, die de inhoud van het video-geheugen karakter per karakter naar de standaard-printer stuurt. De routine GRAPHICS.COM in MS-DOS maakt gebruik van deze routine door eerst te onderzoeken welke video-modus actief is en aan de hand van deze gegevens ofwel de standaard BIOS-routine aan te roepen ofwel een eigen routine in te schakelen, die de printer in de grafische stand zet en nadien de inhoud van het video-geheugen pixel per pixel naar de printer stuurt.
- "Ctrl" + "Alt" + "Break":
  Bij deze combinatie initialiseert de computer door middel van een warme start. Een probaat middel om uit een cash te ontsnappen! Het zal nu echter wel duidelijk zijn waarom dit niet altijd opgaat. De toetsen-combinatie wordt softwarematig gedetecteerd en de processor moet nog wél in staat zijn de BIOS-routine af te handelen die de scan-codes interpreteert. Als de processor echter dol draait in een interne lus komt hij daar niet meer aan toe en dan zal het systeem niet meer reageren op deze magische combinatie.

Dubbele toetsen
Het toetsenbord van de PC heeft een aantal toetsen die schijnbaar dezelfde karakters opwekken. Zo zit het sterretje (*)

zowel op de "8"- als op de "Print Sceen"-toets, zijn er twee hoofdletter-toetsen en komen de cijfers en rekenkundige tekens zowel voor op het hoofd-toetsenbord als op het numerieke bord. Toch hebben al deze toetsen verschillende scan-codes en hoewel de BIOS-routine die de codes leest deze wel omzet in dezelfde ASCII- codes in het lage byte van het karakter-woord onderscheiden deze woorden zich door de scan-code in het hoge byte.
Voor de twee hoofdletter-toetsen zijn zelfs twee bits gereserveerd in het status-byte 417_{hex}, die geset worden als respectievelijk de linker of de rechter "Shift"-ingedrukt wordt.

Het data-formaat van het woord

Inleiding

Zoals reeds geschreven zal de BIOS-routine die zich bezig houdt met het interpreteren van de scan-codes van het toetsenbord, deze omzetten in twee byte brede woorden. Het laagste byte noemt men het hoofd-byte, het hoogste byte het hulp-byte.
Afhankelijk van welke toets werd ingedrukt en van de manier waarop de standaard ASCII-toetsencode wordt gegenereerd, zal de inhoud van deze twee bytes verschillen.
Men onderscheidt drie mogelijkheden, die nu in het kort worden besproken.

Rechtstreekse ASCII-karakters

De ASCII-toetsen zijn de toetsen die een karakter oproepen dat deel uitmaakt van de uitgebreide ASCII-code. Alle letter-, cijfer- en leesteken-toetsen maken hiervan deel uit, maar ook speciale tekens zoals Griekse letters of wiskundige symbolen.
Als op zo'n toets wordt gedrukt, zal het hoofd-byte worden gevuld met de ASCII-code van het karakter en het hulp-byte met de scan-code van het toetsenbord.

"Alt"-methode voor ASCII-karakters

De PC biedt de mogelijkheid ASCII-karakters op een alternatieve manier in te voeren. Stel dat men CHR$(234), het Griekse Ω-teken, wil invoeren. Daar is geen toets voor, dus men moet anders te werk gaan. Men activeert eerst het numerieke toetsenbordje door op de "Num Lock"-toets te drukken. Nadien drukt men de "Alt"-toets in en voert tezelfdertijd via het numerieke toetsenbordje de ASCII-code van het Ω-teken (234) in. Na loslaten van de "Alt"-toets verschijnt het teken op het scherm. Dit is een zeer handige voorziening, omdat het op deze manier mogelijk is bijvoorbeeld kaders rond kleine tekstjes te zetten via een eenvoudige tekstverwerker die niet zélf de mogelijkheid biedt kaders te tekenen. Deze faciliteit wordt rechtstreeks vanuit het BIOS-systeem gestuurd en kan dus vrijwel steeds worden toegepast onder MS-DOS.
Het is op deze manier mogelijk alle ASCII-codes van CHR$(1) tot en met CHR$(255) via het toetsenbord uit te zenden. Het woord van de toets wordt nu echter anders samengesteld. Het hoofd-byte bevat nog steeds de ASCII-code, het hulp-byte blijft echter leeg.
Een volledig overzicht van alle IBM-karakters met hun ASCII-code in decimale en in hexadecimale vorm is gegeven in figuur 3/19.7-6.
De eerste kolom geeft de decimale code, de tweede de hexadecimale notatie, de derde het karakter.

19.8 Principes van de toetsenbord elektronica

0	00		32	20		64	40	@	96	60	`	128	80	Ç	160	A0	á	192	C0	└	224	E0	α
1	01	☺	33	21	!	65	41	A	97	61	a	129	81	ü	161	A1	í	193	C1	┴	225	E1	ß
2	02	☻	34	22	"	66	42	B	98	62	b	130	82	é	162	A2	ó	194	C2	┬	226	E2	Γ
3	03	♥	35	23	#	67	43	C	99	63	c	131	83	â	163	A3	ú	195	C3	├	227	E3	π
4	04	♦	36	24	$	68	44	D	100	64	d	132	84	ä	164	A4	ñ	196	C4	─	228	E4	Σ
5	05	♣	37	25	%	69	45	E	101	65	e	133	85	à	165	A5	Ñ	197	C5	┼	229	E5	σ
6	06	♠	38	26	&	70	46	F	102	66	f	134	86	å	166	A6	ª	198	C6	╞	230	E6	µ
7	07	•	39	27	'	71	47	G	103	67	g	135	87	ç	167	A7	º	199	C7	╟	232	E7	τ
8	08	◘	40	28	(	72	48	H	104	68	h	136	88	ê	168	A8	¿	200	C8	╚	232	E8	Φ
9	09	○	41	29	)	73	49	I	105	69	i	137	89	ë	169	A9	⌐	201	C9	╔	233	E9	Θ
10	0A	◙	42	2A	*	74	4A	J	106	6A	j	138	8A	è	170	AA	¬	202	CA	╩	234	EA	Ω
11	0B	♂	43	2B	+	75	4B	K	107	6B	k	139	8B	ï	171	AB	½	203	CB	╦	235	EB	δ
12	0C	♀	44	2C	,	76	4C	L	108	6C	l	140	8C	î	172	AC	¼	204	CC	╠	236	EC	∞
13	0D	♪	45	2D	-	77	4D	M	109	6D	m	141	8D	ì	173	AD	¡	205	CD	═	237	ED	φ
14	0E	♫	46	2E	.	78	4E	N	110	6E	n	142	8E	Ä	174	AE	«	206	CE	╬	238	EE	ε
15	0F	☼	47	2F	/	79	4F	O	111	6F	o	143	8F	Å	175	AF	»	207	CF	╧	239	EF	∩
16	10	►	48	30	0	80	50	P	112	70	p	144	90	É	176	B0	░	208	D0	╨	240	F0	≡
17	11	◄	49	32	1	81	51	Q	113	71	q	145	91	æ	177	B1	▒	209	D1	╤	241	F1	±
18	12	↕	50	32	2	82	52	R	114	72	r	146	92	Æ	178	B2	▓	210	D2	╥	242	F2	≥
19	13	‼	51	33	3	83	53	S	115	73	s	147	93	ô	179	B3	│	211	D3	╙	243	F3	≤
20	14	¶	52	34	4	84	54	T	116	74	t	148	94	ö	180	B4	┤	212	D4	╘	244	F4	⌠
21	15	§	53	35	5	85	55	U	117	75	u	149	95	ò	181	B5	╡	213	D5	╒	245	F5	⌡
22	16	▬	54	36	6	86	56	V	118	76	v	150	96	û	182	B6	╢	214	D6	╓	246	F6	÷
23	17	↨	55	37	7	87	57	W	119	77	w	151	97	ù	183	B7	╖	215	D7	╫	247	F7	≈
24	18	↑	56	38	8	88	58	X	120	78	x	152	98	ÿ	184	B8	╕	216	D8	╪	248	F8	°
25	19	↓	57	39	9	89	59	Y	121	79	y	153	99	Ö	185	B9	╣	217	D9	┘	249	F9	∙
26	1A	→	58	3A	:	90	5A	Z	122	7A	z	154	9A	Ü	186	BA	║	218	DA	┌	250	FA	·
27	1B	←	59	3B	;	91	5B	[	123	7B	{	155	9B	¢	187	BB	╗	219	DB	█	251	FB	√
28	1C	∟	60	3C	<	92	5C	\	124	7C	¦	156	9C	£	188	BC	╝	220	DC	▄	252	FC	ⁿ
29	1D	↔	61	3D	=	93	5D	]	125	7D	}	157	9D	¥	189	BD	╜	221	DD	▌	253	FD	²
30	1E	▲	62	3E	>	94	5E	^	126	7E	~	158	9E	₧	190	BE	╛	222	DE	▐	254	FE	■
31	1F	▼	63	3F	?	95	5F	_	127	7F	⌂	159	9F	ƒ	191	BF	┐	223	DF	▀	255	FF	

Figuur 3/19.7-6: Een volledig overzicht van alle ASCII-karakters met hun decimale en hexadecimale code.

Niet ASCII-toetsen

Als op een niet ASCII-toets wordt gedrukt, dus op een functie-toets, een cursor-toets, enzoverder of als een toetsen-combinatie met een bepaalde betekenis wordt ingegeven, dan zal het hoofd-byte leeg worden gelaten en het hulp-byte worden gevuld met de scan-code van de toets of een code die uniek is voor de ingedrukte toetsen-combinatie. In figuur 3/19.7-3 werd een overzicht gegeven van de scan-codes van de enkelvoudige toetsen. In de tabel van figuur 3/19.7-7 wordt dit overzicht aangevuld met de codes die worden ingevuld in het byte als men meer dan een toets tegelijkertijd indrukt.

Men spreekt nu niet meer van scan-codes, omdat de scan-codes door het toetsenbord worden gegenereerd en de codes uit de tabel uiteraard door de software uit de opeenvolgende scan-codes worden afgeleid.

19.7 Principes van de toetsenbord elektronica

TOETS-COMBINATIE	CODE	TOETS-COMBINATIE	CODE	TOETS-COMBINATIE	CODE
SHIFT/F1	84	CTRL/F1	94	ALT/F1	104
SHIFT/F2	85	CTRL/F2	95	ALT/F2	105
SHIFT/F3	86	CTRL/F3	96	ALT/F3	106
SHIFT/F4	87	CTRL/F4	97	ALT/F4	107
SHIFT/F5	88	CTRL/F5	98	ALT/F5	108
SHIFT/F6	89	CTRL/F6	99	ALT/F6	109
SHIFT/F7	90	CTRL/F7	100	ALT/F7	110
SHIFT/F8	91	CTRL/F8	101	ALT/F8	111
SHIFT/F9	92	CTRL/F9	102	ALT/F9	112
SHIFT/F10	93	CTRL/F10	103	ALT/F10	113
ALT/A	30	ALT/Q	10	ALT/7	126
ALT/B	48	ALT/R	19	ALT/8	127
ALT/C	46	ALT/S	31	ALT/9	128
ALT/D	32	ALT/T	20	ALT/0	129
ALT/E	18	ALT/U	22	ALT/-	130
ALT/F	33	ALT/V	47	ALT/=	131
ALT/G	34	ALT/W	17	CTRL/PRTSC	114
ALT/H	35	ALT/X	45	CTRL/CURS-L	115
ALT/I	23	ALT/Y	21	CTRL/CURS-R	116
ALT/J	36	ALT/Z	44	CTRL/END	117
ALT/K	37	ALT/1	120	CTRL/PGDN	118
ALT/L	38	ALT/2	121	CTRL/PGUP	132
ALT/M	50	ALT/3	122	CTRL/HOME	119
ALT/N	49	ALT/4	123		
ALT/O	24	ALT/5	124		
ALT/P	25	ALT/6	125		

Figuur 3/19.7-7: De decimale waarde die in het hulp-byte van het toetsenbord wordt gezet als men verschillende toetsen tezelfdertijd indrukt.

Status-bytes

Inleiding

De processor bewaart gegevens die van belang zijn voor de status van het toetsenbord in twee adressen van het geheugen, namelijk in 417_{hex} en in 418_{hex}. Deze adressen wordende de keyboard statusbytes genoemd en het is in principe mogelijk de functie van het toetsenbord vanuit een programma te bepalen door rechtstreeks gegevens in deze twee adressen te schrijven.

Zo kan men bijvoorbeeld de "Caps Lock" inschakelen door het zesde bit van adres 417_{hex} "H" te maken. Dat kan zelfs in BASIC! Voor de actieve software doe-het-zelver mag dus een overzicht van de betekenis van alle bits van deze bytes niet ontbreken!

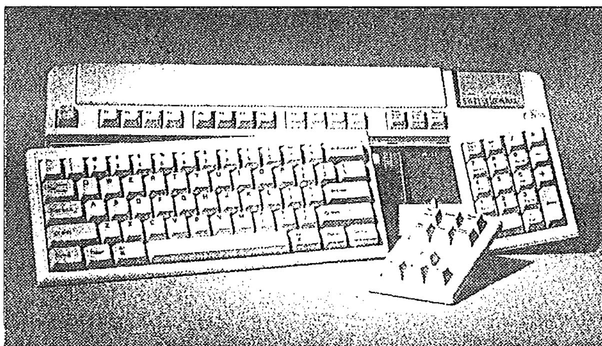
Adres 417_{hex}

De functie van de acht bits van dit adres:

- Bit 0:
  rechtse "Shift"-toets ingedrukt.
- Bit 1:
  linkse "Shift"-toets ingedrukt.

- Bit 2:
  "Ctrl"-toets ingedrukt.
- Bit 4:
  "Scroll Lock"-toets ingedrukt.
- Bit 5:
  "Num Lock" status.
- Bit 6:
  "Caps Lock" status.
- Bit 7:
  "Insert" status.

Een "H" in het bit activeert de functie van het bit. Zo zullen alle letters als hoofdletters geïnterpreteerd worden als men bit 6 "H" maakt. Noteer dat de "Shift"-toetsen in twee bits worden ondergebracht, zodat het in principe mogelijk is te detecteren of een hoofdletter werd opgeroepen door het indrukken van de linker of van de rechter "Shift"-toets.



Figuur 3/19.7-8: Het "programmeerbare" toetsenbord van Datadesk.

Adres 418_{hex}

De bits 4 tot en met 7 van dit byte controleren dezelfde functies als de overeenkomstige bits van adres 417_{hex}. In adres 418_{hex} worden de bits echter alleen "H" als de betreffende toetsen worden ingedrukt. Of deze actie leidt tot het in- of uitschakelen van de betreffende functie wordt in 417_{hex} genoteerd. In feite vormen deze bits van het laagste adres dus een soort software-schakelaars, die telkens als het corresponderende bit in adres 418_{hex} even "H" wordt in- of uitschakelen. De bits 0 tot en met 2 worden niet gebruikt.

Niet standaard toetsenborden

Inleiding

Naast de drie besproken standaard toetsenborden zijn er talloze uitvoeringen op de markt verschenen, die zich door vormgeving of mogelijkheden duidelijk onderscheiden van de standaarden. In dit afsluitend subhoofdstukje worden een paar van de vreemdste uitvoeringen kort besproken.

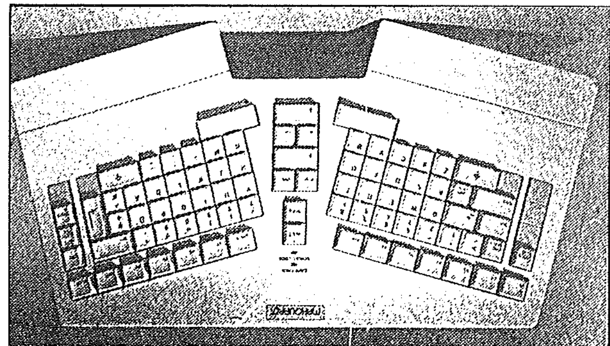
Het Datadesk Switchboard

Dit is een multifunctioneel toetsenbord, hetgeen niet alleen te danken is aan het feit dat het op alle bestaande computersystemen aangesloten kan worden, maar voornamelijk door het feit dat het klavier vrijwel volledig in te delen volgens de eigen smaak. De verschillende toetseneilandjes, zoals de QWERTY-toetsen, het numerieke deel en de cursortoetsen, zijn als losse modules uitgevoerd, zodat deze vrij in de basiseenheid geplaatst kunnen worden. Zo kunnen linkshandige gebruikers/sters besluiten om de cursortoetsen links van de QWERTY-toetsen te plaatsen. In figuur 3/19.7-8 is een impressie gegeven van de manier waarop dat gaat. De drie modules zijn even hoog en kunnen dus in iedere gewenste volgorde in het frame geplaatst worden. De drie eenheden zijn met kabeltjes verbonden met de rest van de elektronica.

19.7 Principes van de toetsenbord elektronica

Ergonomische toetsenborden

Uit onderzoeken is gebleken dat de plaatsing van de toetsen op een standaard toetsenbord nogal wat onnodige verplaatsingen van de vingers en de handen vereist. Diverse fabrikanten hebben naar alternatieven gezocht, die volledig zijn aangepast aan de ergonomie van de mens. Een van deze ergonomische toetsenborden wordt voorgesteld in figuur 3/19.7-9. De positie van de toetsen is een gevolg van een jarenlang onderzoek aan het "Frauenhofer Institut für Arbeitswirtschaft und Organisation" te Stuttgart. Door de strikte scheiding in een linkshandig en een rechtshandig deel zijn dergelijke toetsenborden (die tegenwoordig ook door MicroSoft op de markt worden gebracht) alleen geschikt voor mensen die met tien vingers kunnen typen. Beide helften staan onder een hoek van ongeveer 30 graden en zelfs de afmetingen van de toetsen zijn niet willekeurig gekozen. Uit de genoemde studie bleek dat 19 bij 19 mm^2 de meest ergonomische afmetingen zijn voor een toets. De toetsen zijn dus behoorlijk wat groter dan de toetsen op een standaard toetsenbord en het vervelende aanslaan van twee toetsen tegelijk behoort hiermee dan ook tot het verleden. De onderlinge hoek tussen de twee delen heeft alles te maken met de aanpassing aan het geraamte van het menselijk lichaam. De afstand tussen de schoudergewrichten is immers groter dan die tussen de handen als die dicht bij elkaar op een standaard toetsenbord liggen. Dat is een zeer onnatuurlijke en spanning oproepende houding. Bij de ergonomische toetsenborden wordt dit vermeden door de twee helften zo ten opzichte van elkaar te plaatsen dat de handen in de meest natuurlijke en dus ontspannende houding op het toetsenbord kunnen rusten.



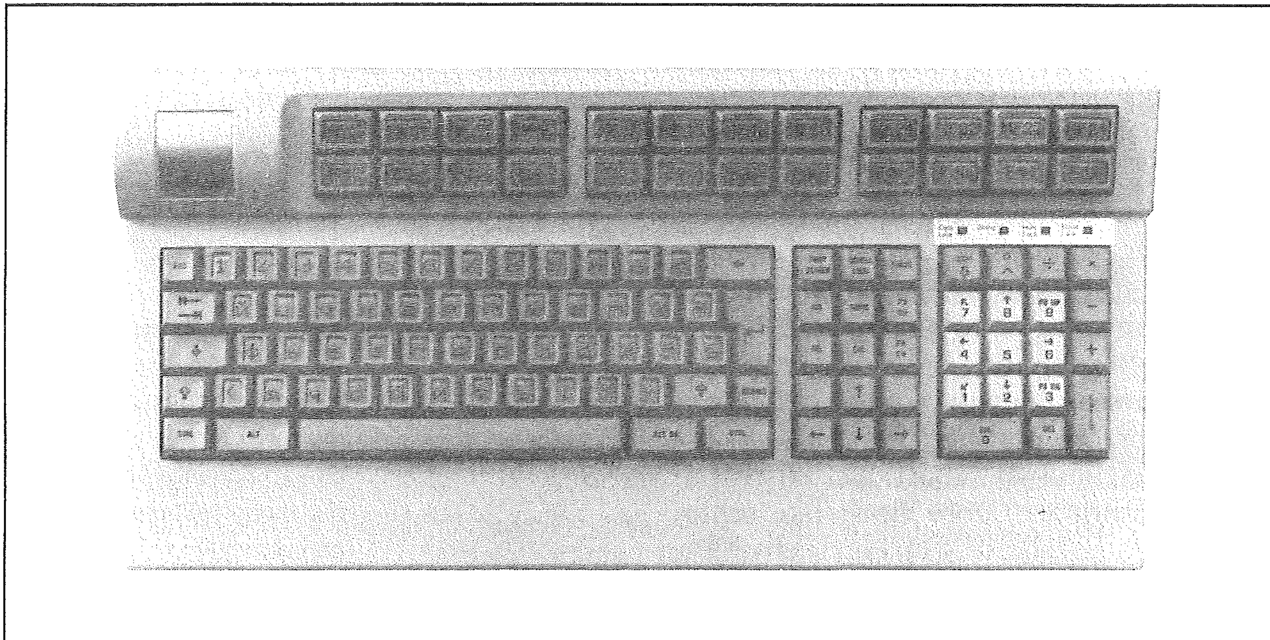
Figuur 3/19.7-9: Bij dergelijke ergonomische toetsenborden is er alles aan gedaan om het typewerk in een zo natuurlijk en ontspannen mogelijke houding te kunnen verrichten.

Een andere eigenschap van dergelijke toetsenborden is dat het numerieke deel niet in het toetsenbord is geïntegreerd, maar als klein los staand extra toetsenbordje is uitgevoerd. Men kan dit bordje dan in de meest gunstige houding op de tafel leggen.

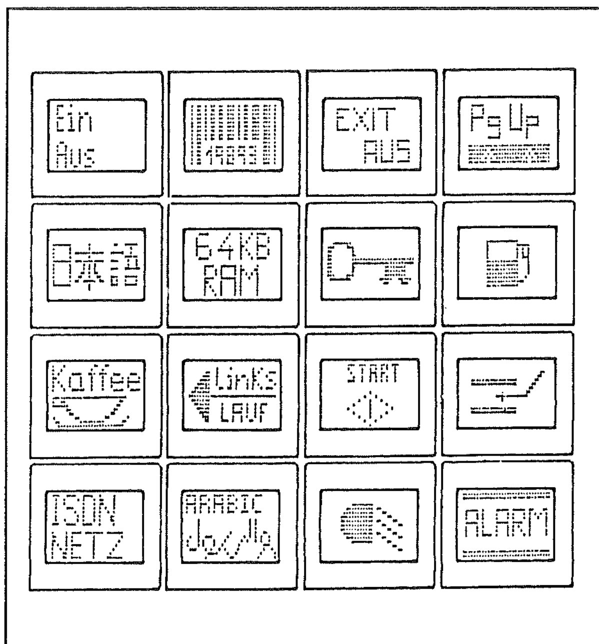
Programmeerbare toetsenborden

Het meest verfijnde toetsenbord is zonder enige twijfel het programmeerbare. Dergelijke apparaten, die overigens meer kosten dan een standaard PC, hebben geen ingegraveerde en ingekleurde toetsopdrukken. Iedere toets bevat een klein alfanumeriek LCD-display, dat individueel te programmeren is. In figuur 3/19.7-10 wordt een dergelijk toetsenbord voorgesteld. Via de bijgeleverde software kan men op een zeer gebruikersvriendelijke manier de "beschrifting" van iedere toets instellen.
Uiteraard kan men niet alleen de letters van het Latijnse schrift programmeren, maar ook Grieks, Cyrillisch, Arabisch, Japans en wetenschappelijke en/of wiskundige symbolen.

19.8 Principes van de toetsenbord elektronica



Figuur 3/19.7-10: Een programmeerbaar toetsenbord.



Figuur 3/19.7-11: Voorbeelden van vrije pictogrammen, die via een grafische editor aan de toetsen kunnen worden toegekend.

De software is hiertoe voorzien van een soort "editor", te vergelijken met de Windows-programmaatjes waarmee men zélf icoontjes kan ontwerpen. Er kunnen 12 complete tekenreeksen in het geheugen van het toetsenbord worden opgeslagen, waarvan er vier direct oproepbaar zijn. Tijdens het werken met een tekstverwerker kan men dus, als men bijvoorbeeld een citaat uit een oud Grieks boek moet schrijven, onmiddellijk overschakelen naar een Grieks toetsenbord!

In figuur 3/19.7-11 zijn een paar zelf ontworpen pictogrammen weergegeven, die men via de editor kan tekenen en nadien aan een toets kan toewijzen. Ieder LCD-schermpje bestaat uit een matrix van 32 bij 16 pixels met ieder een puntgrootte van 0,4 bij 0,6 mm^2. De toetsen zijn voorzien van achtergrondverlichting, zodat het aflezen zelfs in absolute duisternis geen probleem is. Zelfs de kleur van deze achtergrondverlichting is individueel per toets programmeerbaar op rood of groen. Zoals blijkt uit de foto van figuur 3/19.7-10 zijn de functie-toetsen extra groot uitgevoerd en bovendien dubbel.

19.7 Principes van de toetsenbord elektronica

Deze 24 toetsen hebben een extra groot display met een resolutie van 64 bij 16 punten, zodat zelfs zeer ingewikkelde pictogrammen of kleine tekstjes kunnen worden ingegeven. Daarnaast bestaat de mogelijkheid om onder iedere functie-toets een commando-string van maximaal 126 toetsaanslagen te programmeren. Zo kan men dus hardware-macro's opstellen, zoals bijvoorbeeld:

- uitgebreide commando's voor een bepaald software-pakket;
- complexe commando's voor het instellen van printers;
- specifieke teksten en kort standaard zinnen;
- commandoreeksen voor numeriek bestuurde machines.

19.8 Principes van de toetsenbord elektronica

3/19.8

Principes van muizen

Inleiding

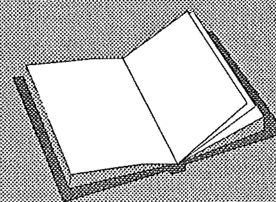
Laat verschijnsel
Het verschijnsel "muis" is voor de PC-gebruiker/ster een vrij nieuw verschijnsel. Terwijl Apple- en Atari-computers al lang standaard werden voorzien van een muis, moest de PC het nog steeds doen met de vier armzalige cursortoetsen. Eerst met het op de markt komen van de zogenoemde "grafische oppervlakteschillen" GEM en Windows (volledig afgekeken van genoemde computers) werd de muis een standaard verschijnsel op de PC.

Soorten muizen
De muis is een standaard attribuut bij de PC geworden en iedere moderne gebruiker/ster werkt er wel eens mee. Er zijn dus een heleboel muizen op de markt gekomen, die in werkingsprincipe in vier grote groepen ingedeeld kunnen worden:
- mechanische muizen;
- optische muizen;
- rugmuizen;
- muistableau's.

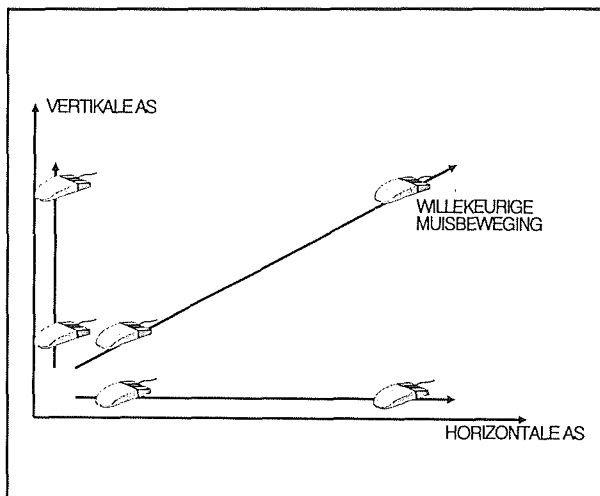
Daarnaast bestaan ook nog eens draadloze muizen die echter niet wezenlijk anders werken dan de vier eerder genoemde soorten, maar waarbij de draadverbinding tussen muis en computer vervangen is door een draadloze infrarode of hoogfrequente link.

Basiswerking
De basiswerking van iedere muis is in wezen identiek. Een willekeurige beweging over een tafel of op een speciale muismat wordt, zie figuur 3/19.8-1, door de elektronica in de muis ontleed in twee bewegingen langs de horizontale en langs de vertikale as. Uiteraard staan deze twee deelbewegingen loodrecht op elkaar. Hierbij evalueert de hardware niet alleen de snelheid van de beweging maar ook de richting.
De muis zendt bij iedere beweging een seriële code, bestaande uit vijf bytes uit. In deze code zit niet alleen informatie over de snelheid en de richting van de muisbeweging, maar ook over het indrukken van een van de muisknoppen. Deze code wordt aangeboden aan de seriële poort van de PC (COM1 of COM2) of via een uitbreidingskaart aan het systeem medegedeeld. De muisdriver die bij het opstarten van het systeem resident in het geheu-

LEES OOK:
Hoofdstuk 3/19.1
Hoofdstuk 6/10.11



gen wordt geladen evalueert de ontvangen codes en vraagt telkens een interrupt bij de processor aan. Via rechtstreekse register-beïnvloeding in de elektronica van de grafische kaart wordt de muiscursor op het scherm gezet. Vandaar dat het kan voorkomen dat een programma muurvast zit, maar dat men de muiscursor nog vrolijk over het scherm kan bewegen. Het is alleen de manier waarop de willekeurige muisbeweging in twee loodrecht op elkaar staande verplaatsingsvectoren wordt omgezet, die verschilt van muistype tot muistype.

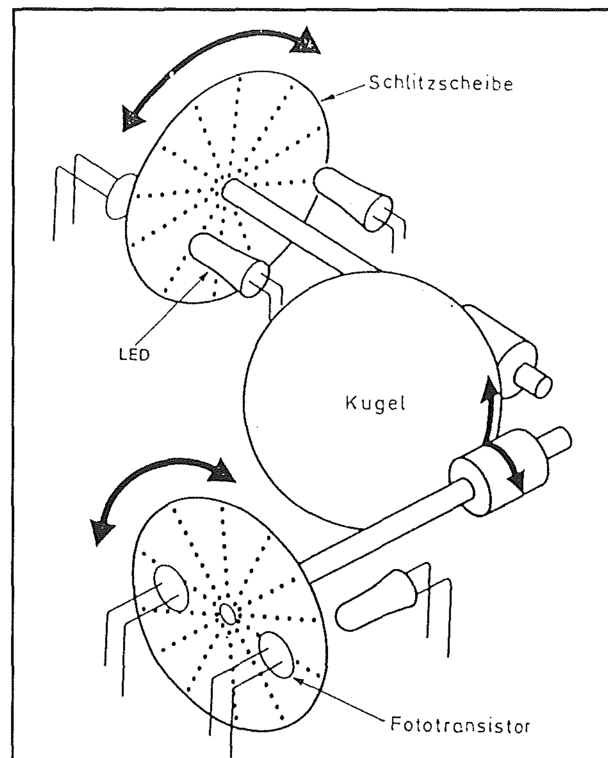


Figuur 3/19.8-1: Het omzetten van de willekeurige muisbeweging in twee verplaatsingsvectoren langs de horizontale en vertikale as.

Mechanische muizen

Constructie

De constructie van een mechanische muis is schematisch voorgesteld in figuur 3/19.8-2. Hart van het apparaat is een vrij grote en zware metalen kogel, die in de meeste gevallen bekleed is met een laagje siliconenrubber. Deze kogel kan vrij bewegen in een bolvormige kooi, waarbij drie met veertjes tegen de kogel gedrukte rubber rolletjes er voor zorgen dat de kogel zo goed als wrijvingsloos in de bolvormige kooi kan draaien. Een van deze rolletjes dient alleen voor het omschreven doel, de twee overige hebben de belangrijke functie de willekeurige beweging van de bol om te zetten in horizontale en vertikale bewegingsvectoren.



Figuur 3/19.8-2: De constructie van een mechanische muis.

De twee rolletjes staan onderling in een loodrechte stand. Als de kogel draait in de met een pijltje aangegeven richting zal het onderste rolletje deze beweging volgen. Het andere rolletje zal echter niet draaien, omdat dit zich precies op de middellijn van de bolbeweging bevindt. Het is alsof de as van de draaiende bol op dat moment door dit andere rolletje gaat. Een

19.8 Principes van muizen

beweging in de getekende richting brengt dus één rolletje aan het draaien, terwijl het andere niet beweegt. Het zal duidelijk zijn dat de rollen omkeren als de beweging van de bol 90 graden draait. Dan zal het achterste rolletje de beweging volgen en het onderste onbeweeglijk blijven. Als de bol een willekeurige beweging gaat uitvoeren, zullen beide rolletjes gaan draaien en wel met een snelheid die evenredig is met de verdraaiïng van de bol in hun richting. Op deze eenvoudige, maar zeer ingenieuze manier wordt de willekeurige beweging van de bol ontleed in twee verplaatsingsvectoren, die loodrecht op elkaar staan.

Van beweging naar pulsen

De bewegingen van de twee loodrecht op elkaar geplaatste assen moeten nu omgezet worden in digitale pulsen, die door het systeem van de computer ingelezen kunnen worden en door de muisdriver geëvalueerd. Er moeten natuurlijk twee verschillende parameters gemeten worden, namelijk:
- de snelheid van de verplaatsing;
- de richting van de verplaatsing.

Het meten van de snelheid

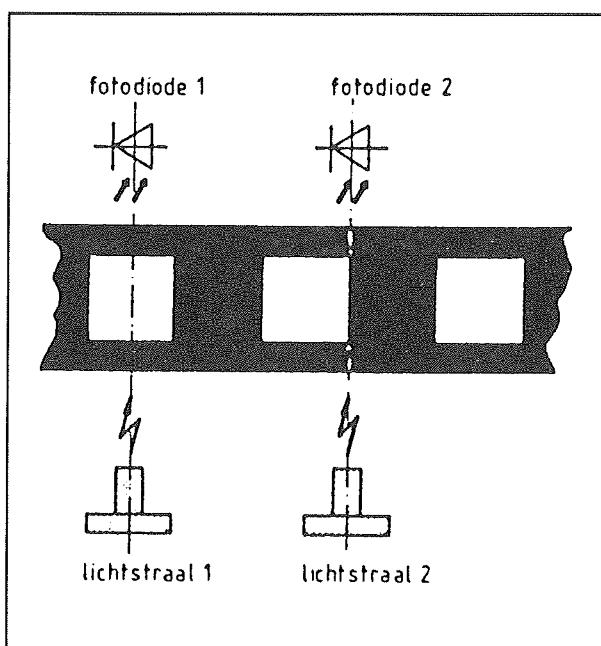
Hoe sneller de kogel draait, hoe sneller ook de twee rolletjes zullen draaien. Het komt er nu op aan deze snelheden te meten en om te zetten in elektrische signalen. Aan de rolletjes zijn assen gekoppeld, aan het uiteinde voorzien van zogenoemde gatenschijven. Dat zijn dunne, zwarte schijfjes, rondom voorzien van een groot aantal kleine gaatjes. Op twee plaatsen draaien deze schijven tussen optische onderbrekingskoppelaars. Zo'n onderdeel bestaat uit een klein kunststof blokje, waarin een gleuf is aangebracht. Aan de ene zijde van de gleuf is een infrarood licht uitstralende diode ondergebracht (LED), aan de andere zijde een optische halfgeleider die gevoelig is voor dat licht. Als er niets in de gleuf aanwezig is valt het licht van de infrarode LED op de lichtgevoelige halfgeleider. Dit onderdeel gaat elektrisch geleiden en zal in de meeste gevallen een "H"-signaal afgeven. Plaatst men iets in de gleuf dat infrarood licht tegen houdt, dan wordt de lichtstraal onderbroken en wordt de lichtgevoelige halfgeleider niet belicht. Dit onderdeel gaat dan sperren, er vloeit geen stroom doorheen hetgeen in de meeste gevallen overeen komt met een "L"-signaal. Als de gatenschijven in de muis tussen de twee optische onderbrekingskoppelaars doordraaien, zal het gatenpatroon in de schijf er voor zorgen dat er afwisselend wél en géén infrarood licht op de fotogevoelige halfgeleider valt. Het gevolg is dat deze voortdurend omschakelt van sperren naar geleiden en een afwisselend "L"- en "H"-signaal afgeeft. Er ontstaat dus een elektrisch pulsvormig signaal op de uitgang. Het zal duidelijk zijn dat het aantal pulsjes per seconde (de frequentie) recht evenredig is met de omtreksnelheid van de schijven. Hoe sneller de schijven draaien, hoe sneller de gaatjes in de schijf de lichtstraal doorlaten en weer verbreken. Op deze manier kan de elektronica bepalen met wat voor snelheid de muis over de tafel wordt bewogen. De verhouding tussen de horizontale en vertikale frequenties kan gebruikt worden om de willekeurige beweging van de muis te reconstrueren op het scherm.

Het meten van de richting

Natuurlijk moet de elektronica ook in staat zijn te detecteren in welke richting de muis over tafel bewogen wordt. Als men alleen zou afgaan op de frequenties

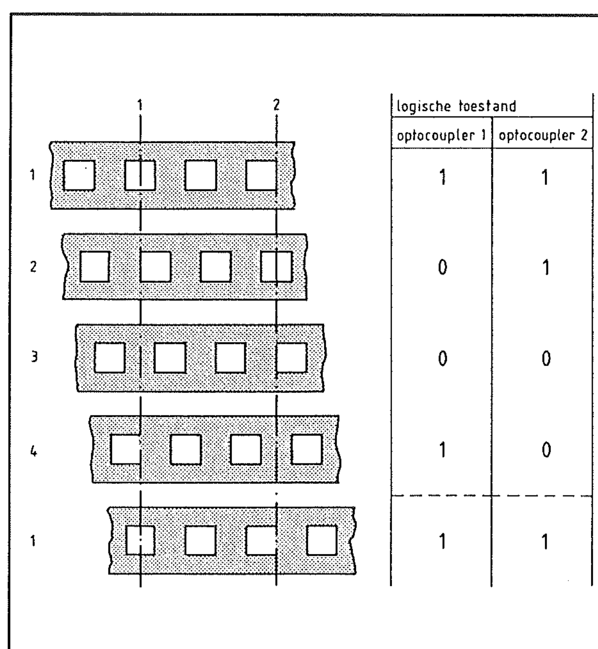
19.8 Principes van muizen

van de twee uitgangssignalen, dan zou de elektronica niet weten of de muis met een bepaalde snelheid van voor naar achter over tafel wordt verplaatst of met dezelfde snelheid van achter naar voor. Om de bewegingsrichting te detecteren wordt gebruik gemaakt van een elektronisch principe dat bekend staat onder de naam "quadratuur-modulatie".



Figuur 3/19.8-3: Het principe van het meten van de bewegingsrichting.

Zoals reeds geschreven draait iedere schijf tussen twee optische onderbrekingskoppelaars. Deze zijn nu zo aangebracht dat als de ene koppelaar precies midden in een gat staat de tweede net op het randje van een ander gat staat. Dit wordt verduidelijkt aan de hand van figuur 3/19.8-3, waarin een segment van de gatenschijf werd "recht gebogen" en werd verondersteld dat de twee optische koppelaars naast elkaar staan. In de praktijk staan deze meestal aan weerszijden van de schijf, maar dat maakt voor het principe niets uit. Door deze specifieke opstelling van de twee koppelaars rond de schijf kan de elektronica heel eenvoudig bepalen in welke richting de schijf draait. In figuur 3/19.8-4 is een voorbeeld getekend, waarbij de schijf (voor de eenvoud weer recht gebogen) van links naar recht wordt bewogen. De logische signalen op de uitgangen van de twee optische koppelaars zijn in de rechter waarheidstabel opgenomen. Er ontstaat nu een typisch bitpatroon, bestaande uit de bitcombinaties "H-H", "L-H", "L-L" en "H-L". Nadien herhaalt deze volgorde zich. De twee uitgangssignalen zijn als het ware 90 graden in fase verschoven. Als de schijf echter van rechts naar links draait, zal een ander bitpatroon ontstaan, namelijk een fasedraaiïng over 90 graden in de andere richting. Het is voor de elektronica (of de software) heel eenvoudig om die twee typische bitpatronen te herkennen.



	logische toestand	
	optocoupler 1	optocoupler 2
1	1	1
2	0	1
3	0	0
4	1	0
1	1	1

Figuur 3/19.8-4: Door de volgorde van de "L"-naar"H"-overgangen in de twee signalen kan de besturingssoftware de richting van de muisbeweging detecteren.

Voor- en nadelen

Mechanische muizen hebben een aantal voor- en nadelen.

Als voordelen zijn te noemen:
- de muis kan over gelijk welk vlak oppervlak bewogen worden, er is geen speciale muismat noodzakelijk;
- hierdoor is het mogelijk met één brede beweging de muiscursor van de linker onderhoek naar de rechter bovenhoek van het scherm te sturen;
- mechanische muizen zijn erg goedkoop en desondanks zeer betrouwbaar;
- alle mechanische muizen, van welk fabrikaat ook, volgen de standaard besturingssoftware van MicroSoft en/of Mouse Systems en kunnen dus zonder problemen in alle muisondersteunende programma's geïnstalleerd worden.

De mechanische muizen hebben echter ook enige nadelen:
- de kogel neemt, vanwege zijn rubber oppervlak, zeer gemakkelijk stof en vuil op van de tafel waarover hij rolt;
- het gevolg is dat deze kogel regelmatig schoongemaakt moet worden;
- doet men dat niet, dan zullen de verontreinigingen zich via de kogel verder verplaatsen in het mechanisme van de muis, waardoor de bol niet meer soepel gaat draaien en de muiscursor met horten en stoten over het beeldscherm beweegt.

Opmerking

De benaming "mechanische muis" is in feite niet correct. De manier waarop de elektrische signalen worden opgewekt is namelijk foto-elektrisch. Men zou de muizen die werken volgens het beschreven principe dus beter "elektro-mechanische" muizen kunnen noemen.

Er zijn immers ook muizen met kogel op de markt, die wél volledig mechanisch werken. De gatenschijven zijn hierbij vervangen door fijntandige tandwielen. De tandwielen rusten tegen de elevatoren van microschakelaars. Als de schijven draaien, zullen de elevatoren van de schakelaars een fractie van een millimeter heen en weer bewegen, als zij over de tandjes van het tandwiel bewegen. Hierdoor worden de micro-schakelaars geopend en gesloten.

Deze constructie, die in de allergoedkoopste muizen wordt toegepast, is echter niet erg betrouwbaar.

Optische muizen

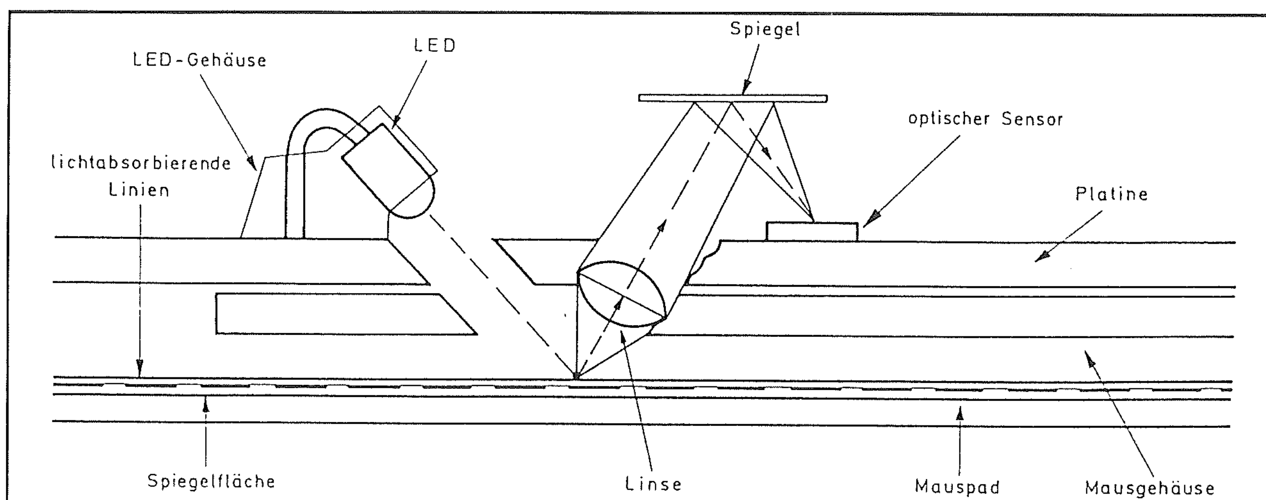
Geen bewegende delen

Optische muizen worden gekenmerkt door het feit dat zij géén bewegende onderdelen bevatten. Zij werken dus volledig volgens elektro-optische principes. Hoewel dit een groot voordeel lijkt (mechanische onderdelen slijten en vervuilen), hebben optische muizen ook een groot nadeel. Deze muizen kunnen niet over een willekeurig plat oppervlak bewogen worden, maar moeten gestuurd worden over een speciale, bij de muis geleverde muismat. Deze speciale matten zijn vrij klein, een lengte van 25 cm en een hoogte van 20 cm is standaard. De muis glijdt soepel over deze muismat, doordat de onderzijde van de muis voorzien is van twee vilten strippen.

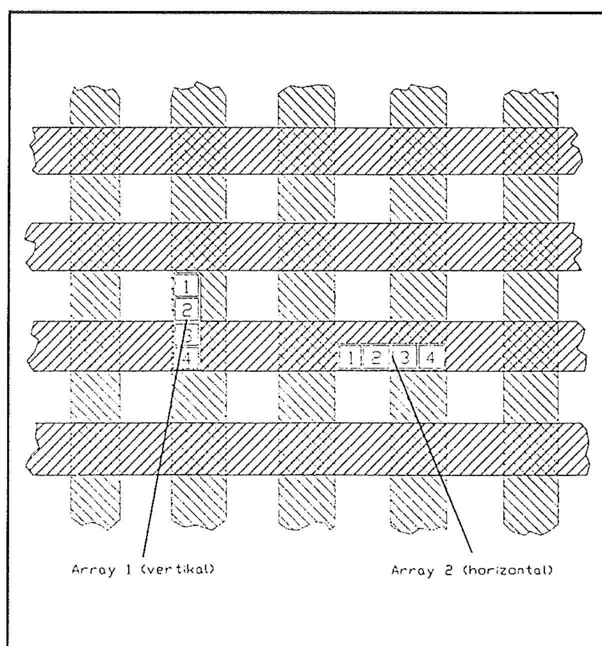
Opmerking

Men kan niet stellen dat er slechts één basisprincipe is, waarmee optische muizen werken. Toch hebben alle toegepaste principes één zaak gemeen. In de muis zitten elektronische onderdelen die licht uitstralen.

19.8 Principes van muizen



Figuur 3/19.8-5: Een doorsnede door een optische muis, waarbij de functie van de voornaamste onderdelen duidelijk herkenbaar is.



Figuur 3/19.8-6: De fundamentele werking van een optische muis: twee optisch gevoelige array's detecteren het licht dat wordt gereflecteerd door de muismat.

Dat licht wordt al dan niet gereflecteerd door het bedrukte oppervlak van de speciale muismat. Het gereflecteerde licht wordt opgevangen door fotogevoelige sensoren. Door het op een speciale manier bedrukken van het oppervlak van de muismat zal het licht afwisselend worden teruggekaatst of geabsorbeerd. De fotodetectoren vangen dit op en aan de hand van deze gegevens kan worden bepaald in welke richting en met welke snelheid de muis over de muismat wordt bewogen.

Het werkingsprincipe

Het werkingsprincipe van een optische muis wordt toegelicht aan de hand van figuur 3/19.8-5. De muismat heeft een reflecterend oppervlak, dat echter bedrukt is met een horizontaal en vertikaal lijnenpatroon, zie figuur 3/19.8-6. De horizontale lijnen absorberen infrarood licht, de vertikale lijnen zichtbaar rood licht. Soms zijn de lijnen vervangen door een raster van puntjes.

De optisch gevoelige sensoren hebben afmetingen die overeen komen met de afstand tussen de lijnen of de puntjes. Bovendien heeft ieder array niet één gevoelig oppervlak, maar diverse. In het voorbeeld van figuur 3/19.8-6 bestaat ieder array in feite uit vier zeer kleine, in een rij opgestelde optische sensoren. De ene sen-

19.8 Principes van muizen

sor, array 1, is gevoelig voor infrarood licht, de tweede (array 2) voor zichtbaar rood licht.
Stel nu dat de muis uit het getekende voorbeeld van onder naar boven wordt bewogen. Sensor-array 1 beweegt over een vertikale lijn op de muismat. Omdat deze lijnen het rode licht absorberen, zal deze sensor alleen de reflectie van het infrarode licht door de horizontale lijnen detecteren. Als de muis van onder naar boven beweegt zal eerst sensor 1 van het array in de door een horizontale lijn opgewekte infrarode schaduw komen. Nadien ondergaan de sensoren 2, 3 en 4 dit lot. In de sensoren worden dus signalen opgewekt. De elektronica kan nu, door deze signalen te evalueren, afleiden dat de muis van onder naar boven wordt bewogen. Bovendien kan de snelheid van de beweging worden berekend. Hetzelfde verhaal geldt uiteraard als men de muis van links naar rechts over de muismat verplaatst. Nu zal array 2, gevoelig voor rood licht, de absorbtie van het rode licht door de vertikale lijnenstructuur detecteren. Deze sensor zal nu vier signalen opwekken, die weer door de elektronica geëvalueerd kunnen worden op snelheid en bewegingsrichting. Als de muis in een willekeurige beweging over de muismat wordt verplaatst, zullen beide sensor-array's signalen genereren. Door het vergelijken van deze signalen kan de elektronica de horizontale en vertikale verplaatsingsvectoren berekenen.

Ook nu quadratuur-modulatie

Door de array's met vier sensoren wordt weer een soort van quadratuur-modulatie toegepast. De faseverschuiving tussen de signalen bepaalt de richting waarin de muis wordt verplaatst. Men kan zich de vraag stellen waarom het noodzakelijk is ieder sensor-array op te bouwen uit vier sensoren. In principe volstaat het twee sensoren toe te passen, net zoals de twee optische koppelaars per schijf bij de mechanische muizen. Dit wordt gedaan om de gevoeligheid van het systeem te verbeteren. Het licht dat door de twee LED's wordt uitgestraald legt een ingewikkelde weg af. Vandaar dat het contrast tussen gereflecteerd licht en geabsorbeerd licht niet erg groot is. Met slechts twee sensoren zou het in de sensoren gegenereerde signaalverschil tussen wél belicht en niet belicht niet erg groot zijn. Als men vier sensoren per array toepast bestaat de mogelijkheid een eenvoudig elektronisch truukje toe te passen om de gevoeligheid van het systeem te vergroten. Men gaat nu het verschil tussen de uitgangssignalen van twee sensoren berekenen en dit verschil vergelijken met het verschil tussen de uitgangssignalen van de twee overige sensoren. Op deze manier neemt de gevoeligheid van het systeem toe en bestaat er een duidelijk onderscheid tussen het uitgangssignaal in belichte en niet belichte toestand.

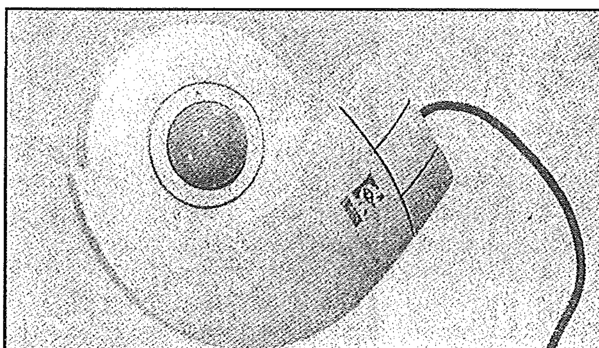
Voor- en nadelen

Ook optische muizen hebben hun typische voor- en nadelen. Het grote voordeel is dat er geen bewegende onderdelen aanwezig zijn die kunnen vervuilen en de werking van de muis verstoren. Maar dat is schijn! Zoals reeds beschreven glijdt de muis over de muismat door middel van twee vilten strippen. Deze strippen nemen het stof over dat op de muismat aanwezig is. Ook andere viezigheden op de muismat worden zonder veel moeite verzameld op de vilten strippen. Na enige maanden zijn de strippen zo smerig, dat zij kleverig worden en de muis niet meer soepel over de muismat glijdt. Soms stelt

men zelfs vast dat het matje, ondanks de rubber bodem, mee met de muis gaat bewegen! Die vilten strippen moeten dus net zo regelmatig schoon gemaakt worden als de kogel van de mechanische muis. Maar dit is een veel eenvoudiger klusje dan het schoonmaken van het interne van een mechanische muis. Een ander voordeel is dat optische muizen een zeer grote nauwkeurigheid hebben en men de muiscursor bijna pixel per pixel over het scherm kan sturen. Vandaar dat dit type muis voornamelijk (ondanks de tamelijk hoge prijs) wordt aangetroffen bij ingenieursbureau's en bij iedereen die zeer nauwkeurig met de computer moet tekenen of ontwerpen.

Rugmuizen

Vaste opstelling
Rugmuizen zijn in feite gewone mechanische muizen, die omgekeerd worden gebruikt. Zoals uit figuur 3/19.8-7 blijkt, zit de kogel bij dit type muizen aan de bovenzijde van het muislichaam.



Figuur 3/19.8-7: Een voorbeeld van een rugmuis, in dit geval van Logitech.

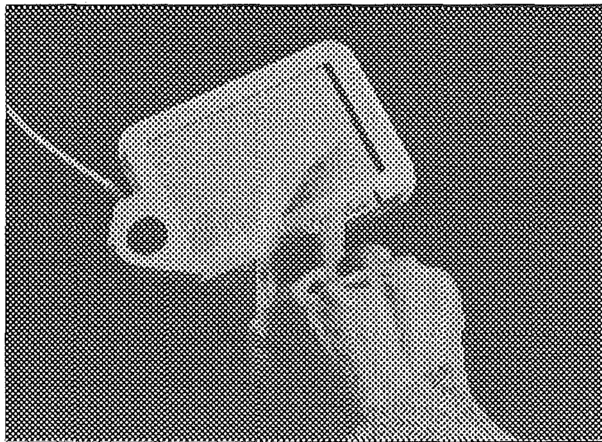
De muis ligt op de tafel en met de palm van de hand of de duim kan men de kogel in alle mogelijke richtingen sturen. In de meeste gevallen is de kogel van een rugmuis vrij groot, zodat de hand een goede grip op het oppervlak heeft en het mogelijk is de kogel heel precies te besturen. Het zal duidelijk zijn dat de elektronische werking van een dergelijke muis volledig te vergelijken is met deze van de mechanische muizen.

Voor- en nadelen
Het grote voordeel van de rugmuis is uiteraard het feit dat men geen vrije ruimte nodig heeft op de werktafel. Vandaar dat dergelijke muizen standaard zijn bij laptop computers, die immers vaak worden gebruikt in tamelijk bekrompen werkomgevingen. Er bestaan zelfs uitvoeringen die men op de laptop kan bevestigen en diverse computers worden tegenwoordig standaard uitgeleverd met een rugmuis die in het toetsenbord is geïntegreerd. Het nadeel van dergelijke muizen is wel dat men moet wennen aan de bediening.

Muistableau's

Miniatuur grafisch tableau
Het muistableau is te vergelijken met de grafische tableau's, die in de professionele ontwerpbureau's worden gebruikt bij het maken van ingewikkelde constructietekeningen middels CAD-software. Zoals uit het voorbeeld van figuur 3/19.8-8 blijkt, bestaan dergelijke muizen uit een klein bordje, waarover men met de top van de vinger moet wrijven. Eventueel kan men ook werken met een stift met stompe punt. Het bordje bevat een matrix van een groot aantal horizontale en vertikale elektrische geleiders, die door de druk van de vinger of de punt van de stift op een bepaalde plaats met elkaar contact ma-

ken. De elektronica tast alle horizontale en vertikale geleiders met een grote snelheid af en kan op een vrij eenvoudige manier detecteren welke horizontale geleider contact maakt met welke vertikale geleider.



Figuur 3/19.8-8: Een voorbeeld van een muistableau, in dit geval de Un-Mouse 2000 van Micro Touch.

Voor- en nadelen

Net zoals de rugmuis heeft het muistableau als voordeel dat er erg weinig ruimte op de werktafel nodig is voor de bediening van de muis. Een nadeel is de vrij hoge prijs van rond de vijfhonderd gulden. Muistableau's hebben echter een heel groot voordeel: men kan de werking zowel relatief als absoluut instellen. In de relatieve modus is de positie van de cursor op het scherm onafhankelijk van de muis. Zet men de vinger ergens op het tableau en verplaatst men de vinger, dan zal de cursorpositie mee verschuiven met de verplaatsing van de vinger. In deze modus werkt het muistableau dus net zoals iedere andere muis. Heeft men de vinger tot aan de rand van het tableau verplaatst en is de muiscursor nog niet op de gewenste plaats, dan tilt men de vinger op, verplaatst hem weer naar de onderste rand van het tableau en verschuift verder. In de absolute modus is er een rechtstreeks verband tussen de plaats van de vinger op het tableau en de plaats van de cursor op het scherm. Links boven op het tableau is links boven op het scherm. Dat is zeer handig bij tekenprogramma's! Men kan bijvoorbeeld een schetsje op het muistableau plakken en dit met de stompe punt van een stift overtrekken. De tekening verschijnt dan keurig op het scherm.

Kalibreren

In de absolute modus moet een muistableau worden gekalibreerd. Dat wil zeggen dat de absolute positie van de vinger op het tableau moet worden afgeregeld op de absolute positie van de muiscursor op het scherm. Of, met andere woorden, als men met de vinger de rechter bovenhoek van het muistableau aanraakt, dan moet de cursor zo ingesteld worden dat deze op dezelfde positie van het scherm staat. Bij alle muizen die volgens dit principe werken wordt hiervoor speciale software meegeleverd.

Draadloze muizen

Geen kabeltje

Draadloze muizen kunnen volgens een van de vier beschreven principes werken. Het enige verschil is dat de draadverbinding tussen muis en PC vervangen wordt door een draadloze verbinding. Het zal duidelijk zijn dat hiervoor flink wat extra elektronica noodzakelijk is, hetgeen zich uit in de prijs.

Draadloze muizen werken volgens twee principes:

- infrarood;
- hoogfrequent.

Infrarood

Bij de infrarode draadloze muizen wordt gewerkt met onzichtbaar, infrarood licht. In de achterkant van de muis, dus die kant die bij normaal gebruik van de muis naar het scherm gericht is, zijn een of meerdere infrarood licht uitstralende LED's aangebracht. Onder, naast of op de monitor wordt een klein kastje bevestigd, waarin een infrarood gevoelige optische sensor zit. Deze sensor pikt het met het muissignaal gemoduleerde infrarood licht op, zet dit weer om in een elektrisch signaal en biedt dit signaal op de normale manier aan een van de seriële poorten aan.

Hoogfrequent

Bij hoogfrequente draadloze muizen bevat de muis een klein radiozendertje. Op een draaggolf van een vrij hoge frequentie wordt het muissignaal gemoduleerd, net zoals muziek bij radio. Ook nu is een kastje nodig, waarin een radio-ontvanger aanwezig is die is afgestemd op de draaggolffrequentie van de muis. Het signaal wordt gedemoduleerd en naar de seriële poort gestuurd.

Voor- en nadelen

Draadloze muizen hebben natuurlijk als groot voordeel dat de soms vervelend over het werkblad slingerende draad afwezig is. Maar naast dit ene voordeel staat een aantal nadelen. Bij de infrarode muizen moet er steeds een optische verbindingsweg aanwezig zijn tussen de muis en de ontvanger. Houdt men de muis scheef, dan zal de ontvanger de infrarode lichtstraal niet oppikken en het systeem zal niet werken. Hoogfrequente muizen hebben van dit verschijnsel geen last. Een tweede nadeel is dat alle draadloze muizen batterijen of accu's aan boord moeten hebben. De elektronica verbruikt tegenwoordig weliswaar erg weinig stroom, maar uiteraard moeten de batterijen regelmatig vervangen worden of de accu's regelmatig opgeladen. Bij accu-gevoede draadloze muizen kan men het opladen verzorgen door een kabeltje aan te brengen tussen de muis en de ontvanger. Ondertussen kan men wel met de muis blijven werken, maar het draadloze karakter is dan natuurlijk weg. In de meeste gevallen kan een draadloze muis ongeveer acht uur werken op één volle accu. Men kan dus het laden verplaatsen naar de nachtelijke uren, zodat men tijdens het werk geen last heeft van de draadverbinding die noodzakelijk is voor het opladen van de accu.
Een handige voorziening bij sommige draadloze muizen is dat één accu in de muis zit en een tweede in de ontvanger. Terwijl de muis werkt op de spanning van de ene accu wordt de tweede opgeladen. Op deze manier kan men dus snel van accu wisselen, zodat men 24 uur per dag met het systeem kan werken.

Eigenschappen van muizen

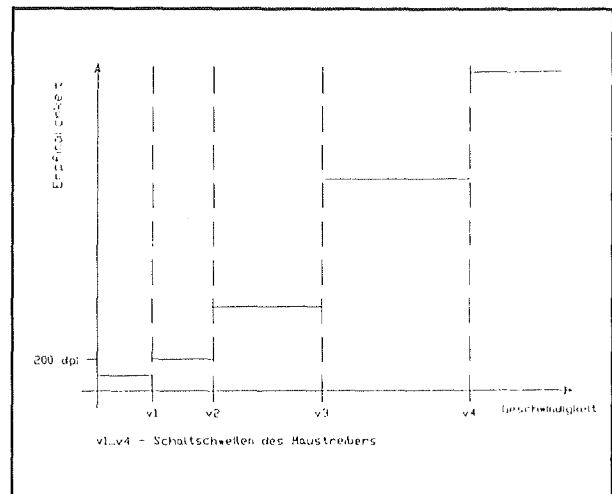
Mickey's

De belangrijkste eigenschap van een muis is de gevoeligheid. Deze parameter definieert het aantal bewegingsimpulsen dat de muis naar de PC stuurt als men de muis één inch in horizontale of vertikale richting verplaatst. De Amerikanen, nooit wars van een grapje, hebben deze parameter de eenheid "Mickey" gegeven. Typische waarden voor de meeste muizen liggen tussen 200 en 400 Mickey's per inch. Dit is de hardware gevoeligheid. Deze gevoeligheid kan softwarematig opgevoerd

worden tot meer dan 10.000 Mickey's per inch.

De gevoeligheid

Aan de gevoeligheid van een muis worden tegenstrijdige eisen gesteld. Enerzijds moet de muis heel gevoelig zijn, om bijvoorbeeld met één beweging de muiscursor van de onderkant van het scherm te verplaatsen naar de menubalk aan de bovenzijde. Maar anderzijds moet de muis vaak extreem laaggevoelig zijn, als men bijvoorbeeld in een CAD-programma het begin van een lijntje heel precies wil laten aansluiten op het eindpunt van een ander lijntje. Vergeet niet dat bij een schermresolutie van 600 bij 800 de muiscursor in principe zonder problemen naar ieder van de in totaal 480.000 pixels op het scherm gestuurd moet kunnen worden!
Het zou dus handig zijn als een muis was voorzien van een gevoeligheidsschakelaar, waarmee men de gevoeligheid kan omschakelen tussen bijvoorbeeld 200 en 5.000 Mickey's per inch. Toch zal men maar weinig muizen aantreffen die een dergelijke handige voorziening hebben.
In de meeste gevallen wordt deze omschakeling van de gevoeligheid dynamisch geregeld door de bij de muis geleverde software. Tegenwoordig wordt meestal gebruik gemaakt van een niet-lineaire dynamische gevoeligheidsregeling. Als men de muis snel beweegt, dan zorgt de software ervoor dat de gevoeligheid van de muis zeer hoog wordt ingesteld.
Beweegt men de muis traag, dan zal de software de gevoeligheid terug schroeven. In de meeste gevallen wordt gewerkt met drie tot vier zogenoemde "drempels", waardoor de niet-lineaire dynamische gevoeligheidsregeling ontstaat. Dit wordt toegelicht aan de hand van de grafiek in figuur 3/19.8-9.



Figuur 3/19.8-9: De karakteristiek van de niet-lineaire gevoeligheidsregeling van moderne muissoftware.

Op de horizontale as wordt de bewegingssnelheid van de muis uitgezet. De vertikale as geeft de gevoeligheid van de muis in Mickey's per inch. Bij een snelheid die lager is dan v_1 stelt de software de gevoeligheid in op 100 Mickey's per inch. Bij een snelheid tussen v_1 en v_2 wordt de gevoeligheid verhoogd tot 200 Mickey's per inch, etc. De vier snelheden v_1, v_2, v_3 en v_4 zijn de vier drempels van de karakteristiek.

Van muis naar PC

Serieel maar niet gestandaardiseerd

De meeste muizen worden tegenwoordig geleverd als serieel invoer apparaat. Dat betekent dat de muis met een 9- of 25-polige D-connector op een van de seriële COM-poorten van de PC moet worden aangesloten. Toch is de muis geen echt serieel apparaat. Het door de muis gehanteerde protocol voldoet namelijk niet aan de RS-232 standaard, waarmee de COM-poorten werken. Vandaar dat een speciale

muisdriver moet worden geladen, die op register-niveau de gebruikte COM-poort aanpast aan het speciale muisprotocol.
De door de muis verstuurde signalen worden alleen aangeboden aan de RS-232 lijn RD. Deze bevindt zich op pen 2 van de 9-pens connector en op pen 3 van de 25-pens connector. Alle overige lijnen zijn uitgeschakeld en bevinden zich op laag niveau. Dit lage niveau komt bij de RS-232 interface echter overeen met een gelijkspanning van -12 V. Van dit gegeven wordt dankbaar gebruik gemaakt! Deze spanning wordt in de muis gebruikt als negatieve voedingsspanning voor de elektronica.

De drukknoppen

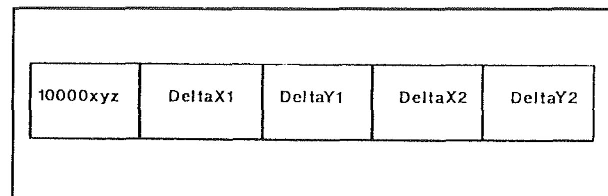
Iedere muis is uitgerust met twee of drie drukknoppen. Het zal duidelijk zijn dat de muis niet alleen signalen naar de COM-poort moet versturen over de horizontale en vertikale verplaatsing, maar ook over de status van de drukknoppen.

Het muissignaal

De muis communiceert over slechts één lijn met de COM-poort en het zal dus zonder meer duidelijk zijn dat de signalen in een seriële vorm verstuurd moeten worden.
Dat gebeurt onder de vorm van een woord dat uit vijf bytes bestaat. Dit woord wordt uitgezonden als:
- een van de drukknoppen wordt ingedrukt of losgelaten;
- zolang de muis verplaatst wordt.

Deze signalen wekken, via de muisdriver, interrupts INT33h op die het plaatsen van de muiscursor op het scherm besturen.
Ieder van de vijf tekens bestaat uit:
- een startbit;
- 8 databits (1 byte);
- 1 stopbit.

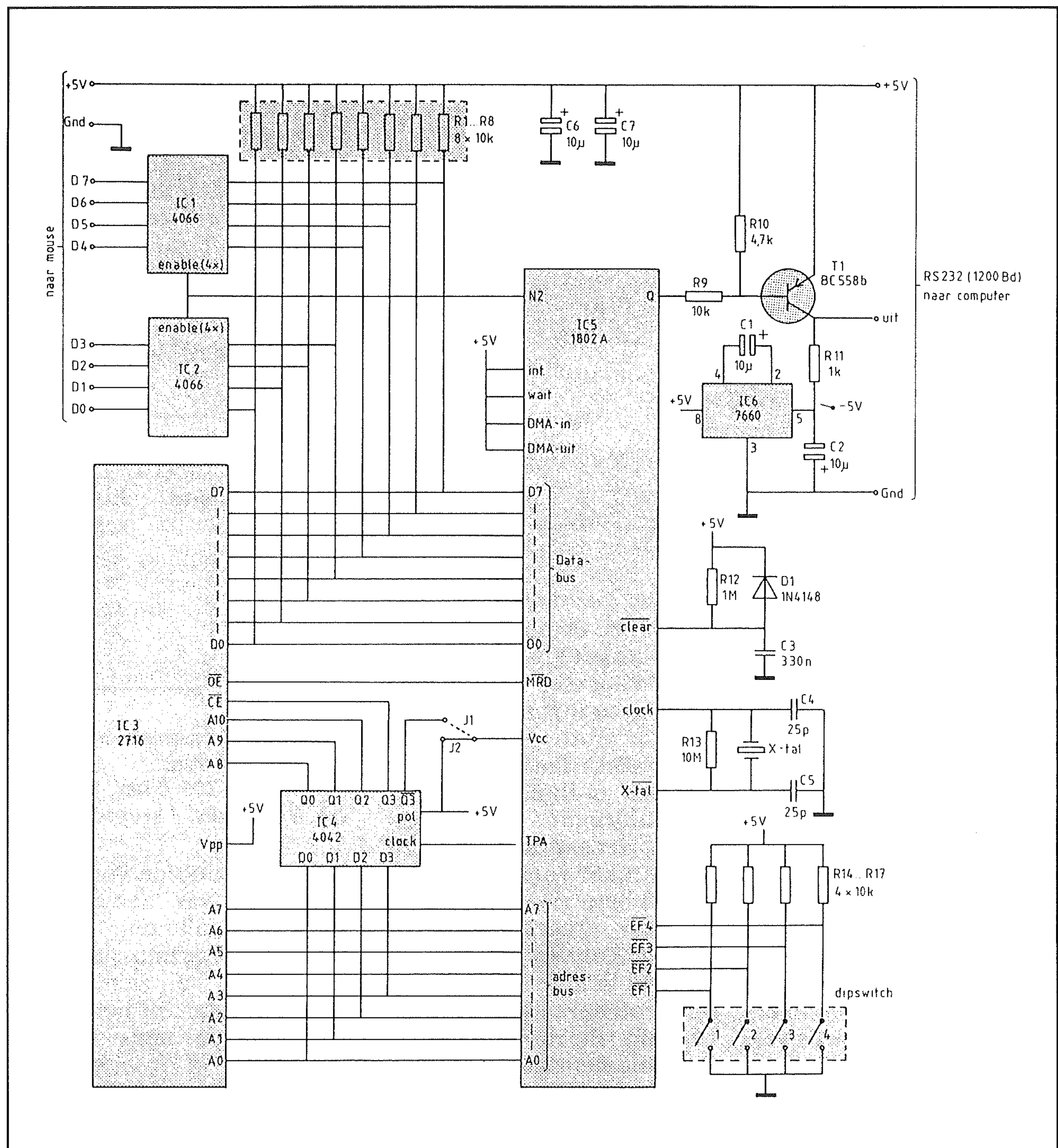
Er wordt geen pariteit toegepast, de snelheid is gestandaardiseerd op 1.200 bit/s. De samenstelling van de vijf bytes is geschetst in figuur 3/19.8-10.



Figuur 3/19.8-10: De samenstelling van één muiswoord.

De vijf eerste bits van het eerste byte van het woord bevatten steeds de code "H-L-L-L-L". Deze code kan door de software gedetecteerd worden als begin van een muiswoord. De drie laatste bits X, Y en X definiëren de status van de drie (of twee) drukknoppen. Een ingedrukte toets komt overeen met een "L". X geeft de status van de linker toets, Y deze van de middelste toets en Z deze van de rechter toets. Nadien volgen vier bytes DeltaX1, DeltaY1, DeltaX2 en DeltaY2, die de positieverandering van de muis weergeven sinds het vorige woord werd verstuurd. X registreert de horizontale bewegingen, Y de vertikale. De waarden van de Delta's kan liggen tussen -128 en +127. De positieve getallen worden in binaire code uitgezonden, de negatieve in het 2-complement formaat. Positieve getallen staan voor bewegingen van links naar rechts respectievelijk van onder naar boven. De bewegingen worden genoteerd in honderdsten van een inch, de eenheid van X en Y is dus omgerekend 0,254 mm.
Als de muis niet wordt bewogen, maar alleen een knop wordt geactiveerd, bevatten de Delta-bytes de decimale code 00. Het uitgezonden woord bevat dan bijvoorbeeld de decimale codes 83-00-00-00-00.

19.8 Principes van muizen



Figuur 3/19.8-11: Het schema van een muis-interface rond een 1802 microprocessor.

De elektronica van de muis

Bij het verwerken van de door de sensoren geleverde pulstreinen naar een serieel signaal komt nogal wat kijken. Tijdens het uitzenden van een muiswoord kan het natuurlijk gebeuren dat de muis in beweging is. Deze bewegingen moeten worden gedetecteerd, opgeslagen in een geheugen en bij de eerste de beste gelegenheid naar de PC verzonden. Het zal duidelijk

19.8 Principes van muizen

zijn dat dit een klusje is dat het beste door een kleine microprocessor kan worden geklaard. Deze processor moet zich dan zowel bezig houden met het detecteren van bewegingen, knopdrukken als met het uitzenden van de seriële code. Dat is echter niet zo'n groot probleem. Bij het uitzenden van de (trage)seriële code moet men wachtlussen inbouwen. Deze wachtlussen kunnen vervangen worden door de routines voor het inlezen van de pulsen die door de elektronica van de muis worden geleverd.
In figuur 3/19.8-11 is, ter illustratie, het schema getekend van een door "Radio Bulletin" ontwikkelde muis-interface. Hart van de schakeling is een processor van het type 1802. Bij het opkomen van de voedingsspanning zorgt het netwerkje R12, D1 en C3 ervoor dat de processor gereset wordt. Het besturingsprogramma is opgenomen in een EPROM van het type 2716 (IC3). De van de muis-elektronica afkomstige pulsen worden via acht elektronische schakelaars IC1 en IC2 aangeboden aan de data-bus van de processor. Op deze heel eenvoudige manier wordt gegarandeerd dat het inlezen van de muispulsen niet in conflict komt met het data-verkeer tussen de processor en het programma in EPROM. De acht ingangen zijn als volgt met de muis te verbinden:
- D0: linker muisknop;
- D1: middelste muisknop;
- D2: rechter muisknop;
- D3: niet in gebruik;
- D4: X1 richtingspulsen;
- D5: X2 richtingspulsen;
- D6: Y1 richtingspulsen;
- D7: Y2 richtingspulsen.

Met de twee DIP-schakelaars S1 en S2 kan men de communicatiesnelheid instellen:
- S1 UIT, S2 UIT: 600 Bd;
- S1 UIT, S2 AAN: 1.200 Bd;
- S1 AAN, S2 UIT: 1.200 Bd;
- S1 AAN, S2 AAN: 2.400 Bd.

3/19.9

Principes van de beeldscherm elektronica

Inleiding

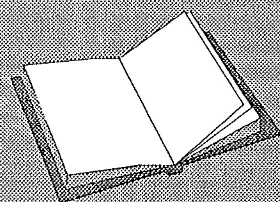
Snelle evolutie
Als er één terrein van de computer elektronica is, waar de ontwikkelingen werkelijk razendsnel zijn gegaan, dan is het wel dat van de beeldbuis besturing. In amper twaalf jaar is de originele MDA, ontwikkeld door IBM, opgevolgd door een aantal standaarden, die steeds betere weergavekwaliteit boden en bieden. Op dit moment zijn grafische kaarten ware wondertjes der techniek, die intelligente processoren bevatten en communiceren via een eigen AGP-bus met de processor. Van MDA tot deze AGP: in dit hoofdstuk wordt een volledig overzicht gegeven van deze schitterende evolutie, waarvan de grenzen nog lang niet in zicht zijn!

Standaarden
Wil een PC-systeem zonder problemen de resultaten van programma's op het scherm zetten, dan is het absoluut noodzakelijk dat er afspraken worden gemaakt over de manier waarop dat gebeurt. Vandaar dat de beeldscherm elektronica, zolang de PC bestaat, beheerst wordt door een aantal kreten, die steeds weer een nieuwe standaard beschreven. Iedere nieuwe standaard moest per definitie compatibel zijn met alle vorige standaarden, want anders zou het kunnen gebeuren dat de trotse bezitter van een nieuwe grafische kaart merkt dat oude programma's opeens niet meer willen werken!

Overzicht
Als men de ontstaansgeschiedenis van de moderne PC doorloopt, komt men op het gebied van de beeldbuis elektronica de onderstaande standaarden tegen:
- MDA;
- Hercules;
- CGA en MCGA;
- EGA en PGC;
- VGA;
- EVGA;
- SVGA;
- XGA.

Nadien ontstaat een periode van grote onrust. Niet toevallig valt deze periode samen met de sterke opkomst van MicroSoft's Windows. Dit volledig grafisch georiënteerd besturingssysteem stelt immers zeer hoge eisen aan de aansturings-

LEES OOK:
Hoofdstuk 3/19.1
Hoofdstuk 3/19.2
Hoofdstuk 5/3.1.4
Hoofdstuk 5/3.1.5
Hoofdstuk 5/3.1.6
Hoofdstuk 6/10.3



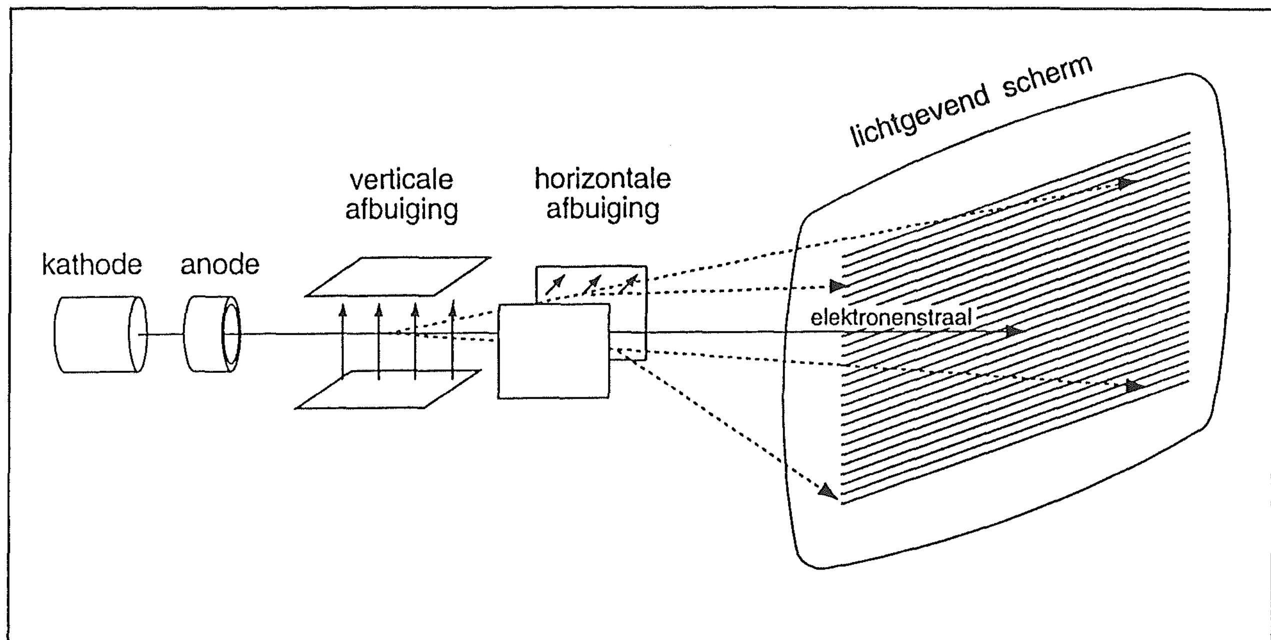
elektronica van het beeldscherm. Maar bovendien werkt Windows met een systeem, waarbij perifere apparaten toegang krijgen tot gestandaardiseerde aansturingsprogramma's die in Windows geïntegreerd zijn. Alle Windows-programma's die iets naar het beeldscherm willen schrijven, roepen deze programma's aan. Deze programma's roepen nadien bepaalde registers aan waarmee bepaalde functies van de grafische kaart worden geactiveerd. Dat geeft fabrikanten van grafische kaarten de mogelijkheid uitgebreid te experimenteren met nieuwe, snelle hardware. Het volstaat immers dat de drivers van de grafische kaart op systeem-niveau kunnen samenwerken met de universele drivers die in Windows zitten.
Er werden tientallen zogenoemde "accelerators" ontwikkeld: gespecialiseerde grafische processoren die in staat waren bepaalde grafische instructies volledig zelfstandig uit te werken, zodat de processor van de PC verlost werd van een heleboel rekenwerk. Vanaf dat moment kan men stellen dat er van een echte standaard in de elektronica van de grafische kaarten geen sprake meer is. De standaardisatie zit hem in de laag die Windows automatisch in de besturingssoftware inbouwt en waarmee de eigen firmware van de grafische kaarten moet kunnen communiceren.

Video en 3D
Na de periode van de "accelerator"-kaarten verschijnt een geheel nieuwe uitdaging aan de horizon. Er wordt software ontwikkeld die driedimensionele beelden op de monitor kan zetten en die in staat is video zodanig te comprimeren, dat bewegende beelden met een snelheid van 25 frames per seconde van de harde schijf naar het beeldscherm gestuurd kunnen worden. Beide nieuwe mogelijkheden stellen opnieuw immens hoge eisen aan de nieuw te ontwikkelen grafische kaarten. Zowel het in real time opbouwen van driedimensionele beelden (het zogenoemde renderen) als het weergeven van die gecomprimeerde video is alleen mogelijk als volledig nieuwe systemen worden ontwikkeld, zoals supersnelle digitaal naar analoog omzetters, grafische processoren die met een busbreedte van 128 bit werken, speciale snelle geheugens op de grafische kaarten en nieuwe bussystemen, die de grafische kaarten op de snelst mogelijke manier verbinden met de rest van het systeem.
Deze evolutie is op dit moment in volle gang en echte standaarden zijn nog niet gezet. Misschien dat de aan het eind van dit hoofdstuk besproken AGP wel die nieuwe standaard zal worden!

Principe van beeldschermbesturing

De opbouw van het beeld
In principe werken PC-monitoren in grote lijnen volgens hetzelfde principe waarmee TV's werken.
De inhoud van een beeld wordt verdeeld over een aantal lijnen, die onder elkaar op het beeldscherm worden geschreven, zie figuur 3/19.9-1. Iedere lijn is weer opgebouwd uit een aantal beeldpunten of *pixels*, die naast elkaar op de lijn worden geschreven. Nadat één volledig beeld is geschreven, begint het proces opnieuw. Door de traagheid van het menselijke oog lijkt het of er een flikkervrij beeld op het scherm staat.

19.9 Principes van de beeldscherm elektronica



Figuur 3/19.9-1: Het principe van de beeldopbouw bij PC-monitoren.

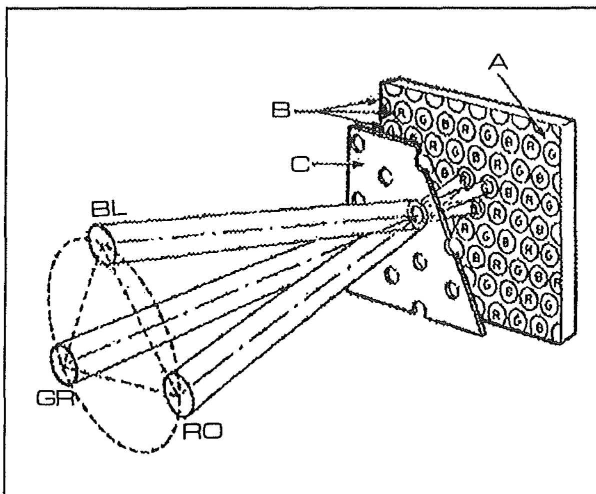
In tegenstelling tot het Europees gestandaardiseerde TV-systeem, ligt het aantal lijnen dat wordt geschreven, het aantal pixels per lijn en het aantal beelden per seconde niet vast. Deze drie gegevens worden volledig bepaald door de elektronica van de grafische kaart die in de PC is geïnstalleerd.

Resolutie
Een belangrijk gegeven van iedere grafische kaart is dan ook de maximale resolutie die de kaart kan genereren. Twee getallen die respectievelijk het aantal lijnen per scherm en het aantal pixels per lijn definiëren. Een VGA-kaart met een resolutie van 600 bij 800 bouwt een beeld dus op uit 600 lijnen, die ieder 800 pixels bevatten. Moderne grafische kaarten hebben vaak een maximale resolutie van 1.600 bij 1.200, waardoor beelden op het scherm worden opgebouwd van een fotorealistische kwaliteit.

Vertical refresh-rate
Een tweede belangrijke eigenschap van iedere grafische kaart is de maximale vertical refresh-rate. Dit is een getal, dat aangeeft hoe veel volledige beelden er per seconde worden geschreven. Bij TV is dat standaard 50, bij grafische kaarten van uitzonderlijke kwaliteit kan dit getal oplopen tot meer dan 100. Het zal duidelijk zijn dat een hogere refresh-rate een veel rustiger beeld oplevert!

Kleurendiepte
Grafische kaarten werken per definitie digitaal. Dat wil zeggen dat de informatie over de kleur van ieder pixel ook digitaal wordt opgeslagen en bewerkt. Bij kleurenmonitoren wordt het systeem van additieve kleurmenging toegepast. Dat wil zeggen dat het scherm is bedrukt met dotjes van drie verschillende fluorescerende stoffen, die bij bestraling met elektronen respectievelijk rood, groen en blauw licht

uitstralen. De grafische kaart genereert drie signalen, die ieder een van de drie elektronenstralen GR, BL en RO besturen, zie figuur 3/19.9-2. Bij de eerste generaties kaarten werden de drie elektronenstralen digitaal aangestuurd. Het gevolg was dat deze kaarten maar acht kleuren konden genereren, namelijk de drie basiskleuren van het scherm plus hun additieve mengkleuren wit, magenta, geel en natuurlijk zwart, zie figuur 3/19.9-3.



Figuur 3/19.9-2: Het genereren van kleuren door het aansturen van de fluorescerende dots rood, groen en blauw door middel van drie elektronenstralen.

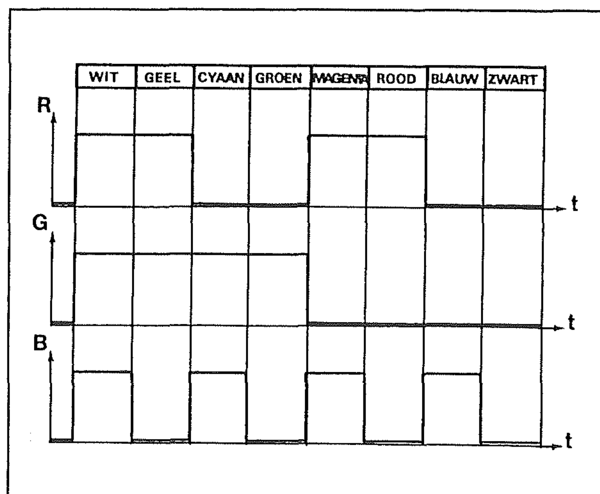
Het zal duidelijk zijn dat acht kleuren uit de tijd is. Vandaar dat de kleurendiepte een derde belangrijke eigenschap is van grafische kaarten. Met dit getal wordt aangegeven hoeveel verschillende kleuren de kaart kan genereren door het in bepaalde verhoudingen mengen van de drie basiskleuren. Moderne grafische kaarten zijn dan in staat meer dan 16 miljoen kleuren op het beeldscherm te zetten. Toch wordt de informatie over al deze kleuren digitaal verwerkt. Dat kan alleen als er veel meer dan één bit per basiskleur beschikbaar is.

Voor meer dan 16 miljoen kleuren heeft men in totaal per pixel 24 bit nodig: 8 voor rood, 8 voor groen en 8 voor blauw. Deze bits worden in een snelle digitaal naar analoog omzetter omgevormd in een analoge spanning, die de intensiteit van de elektronenstraal regelt. Men heeft dan 256 intensiteiten per kleur, hetgeen in totaal 256 x 256 x 256 = 16.777.216 kleurentinten oplevert.

Er zijn een aantal standaard kleurendiepten ontstaan:

- true color:
  24 bit, 16.777.216 kleuren;
- high color:
  16 bit, 65.536 kleuren;
- low color:
  8 bit, 256 kleuren.

Bij de 16 en 8 bit systemen krijgt niet iedere kleur even veel bits toegewezen.



Figuur 3/19.9-3: Het ontstaan van mengkleuren door de additieve menging van de drie basiskleuren.

De pixel clock rate

Het zal duidelijk zijn dat de tijd die ter beschikking staat voor het schrijven van één dot afhankelijk is van de resolutie en van de vertical refresh rate. Hoe meer

punten, lijnen en beelden er per seconde moeten worden geschreven, hoe minder tijd er is voor het schrijven van één punt. Dat heeft consequenties voor de klokfrequentie waarmee de drie digitaal naar analoog omzetters moeten werken. Dit voert tot een vierde belangrijk gegeven van grafische kaarten, de pixel clock rate. Dit is de frequentie waarmee de DAC's gestuurd worden en waarmee de drie digitale gegevensstromen ROOD, BLAUW en GROEN in analoge spanningen worden omgezet. Moderne grafische kaarten hebben een pixel refresh rate die ver boven 100 MHz ligt. Wie de gegevens van "normale" DAC's in de data-handboeken even doorloopt, zal tot de conclusie komen dat dit een immens hoge snelheid is. Vandaar dat dergelijke omzettingssnelheden alleen maar gehaald kunnen worden door speciale schakelingen te ontwerpen, die video-DAC's worden genoemd.

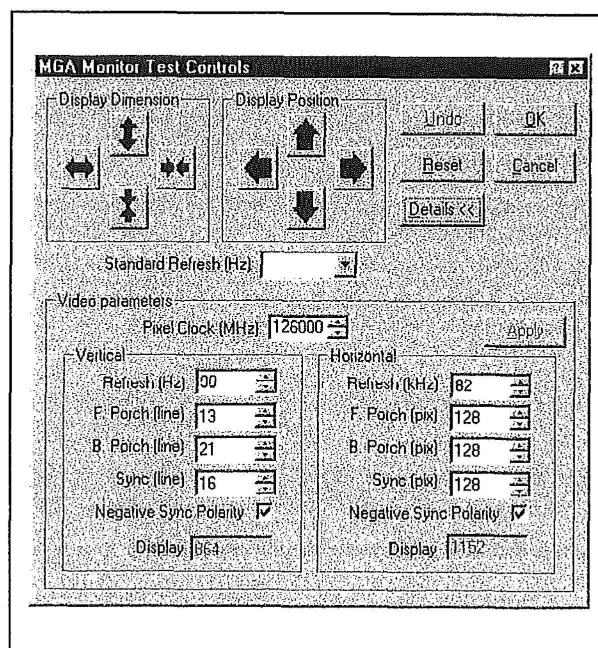
Video-BIOS

BIOS is het letterwoord voor "Basic Input/Output System". Het BIOS van een computer is het meest fundamentele besturingssysteem, dat vast in een EPROM is ingeprogrammeerd en dat er voor zorgt dat het systeem, na het inschakelen van de voedingsspanning, de meest primaire activiteiten kan uitvoeren. De meeste videokaarten hebben een eigen BIOS nodig, dat er voor moet zorgen dat het systeem onmiddellijk in staat is gegevens naar het scherm te sturen, ook als er nog geen besturingssoftware of grafische drivers van de harde schijf zijn geladen.
Dat BIOS zit geprogrammeerd in EPROM's die op de grafische kaart zélf aanwezig zijn. De adressen van deze geheugens worden gespiegeld naar het systeemgeheugen, zodat er steeds een geheugen-segment is dat men niet kan gebruiken, omdat dit segment dezelfde adressen heeft als de BIOS-chip's van de grafische kaart.

Samenvatting

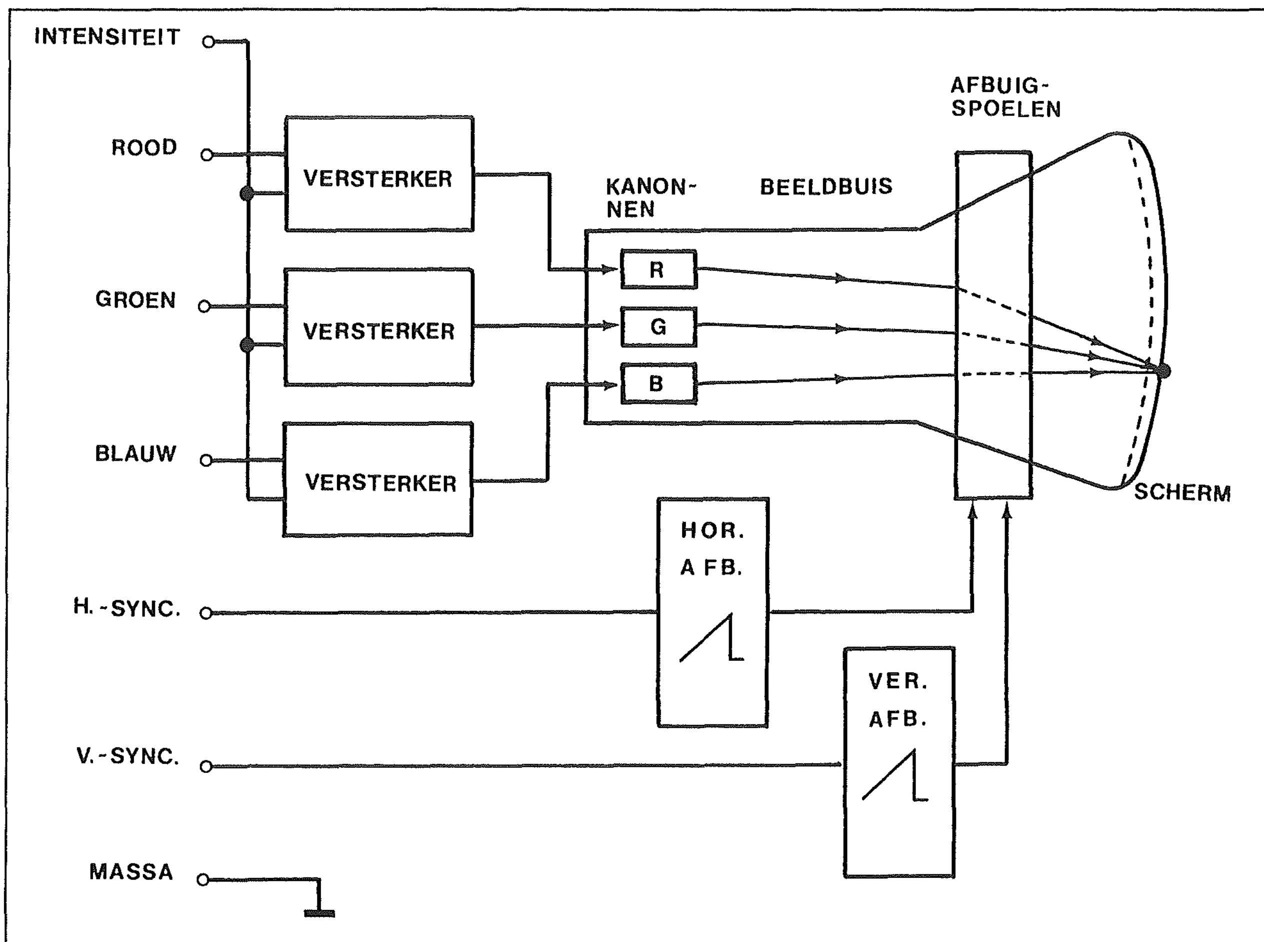
Grafische kaarten worden gedefinieerd door vier belangrijke eigenschappen:
- de resolutie;
- de vertical refresh-rate;
- de kleurendiepte;
- de pixel clock rate.

In de meeste gevallen wordt er bij iedere moderne grafische kaart een programma geleverd, waarmee men al deze eigenschappen tot in de details kan instellen. Als voorbeeld is in figuur 3/19.9-4 het instelschermpje getoond, dat bij de vrij moderne grafische kaart Millenium van Matrox wordt geleverd.



Figuur 3/19.9-4: Met dit programma "MGA Monitor Test Controls" kan men de belangrijkste specificaties van de Millenium grafische kaart instellen.

19.9 Principes van de beeldscherm elektronica



Figuur 3/19.9-5: Het algemeen blokschema van een kleurenmonitor.

Blokschema van een kleurenmonitor

In figuur 3/19.9-5 is het algemene blokschema van een kleurenmonitor getekend. Drie versterkers ontvangen de drie *gescheiden* signalen die de rode, groene en blauwe pixels aansturen. Vroeger was nog een extra intensiteit-signaal aanwezig, omdat sommige oude standaarden van acht zestien kleuren wisten te maken door de intensiteit van de drie elektronenstralen gezamenlijk te verhogen. Naast hel rood had men dan de beschikking over rose. De drie versterkers sturen uiteindelijk de drie elektronenkanonnen op een manier die volledig vergelijkbaar is met deze die bij kleuren-TV wordt toegepast. De synchronisatiesignalen die de elektronenstralen weer naar links sturen na het schrijven van een lijn en weer naar boven na het schrijven van een beeld worden ook afzonderlijk aan de monitor aangeboden.

Een PC-monitor heeft dus minstens zes signalen nodig:

- ROOD;
- GROEN;
- BLAUW;
- H-SYNC;
- V-SYNC;
- Massa.

19.9 Principes van de beeldscherm elektronica

MDA

Inleiding
MDA is het letterwoord van "Monochrome Display Adapter". De allereerste PC's werden met deze door IBM ontwikkelde standaard uitgerust. Kenmerk van de MDA is dat hij alleen in staat was tot het weergeven van karakters. Grafische afbeeldingen weergeven was er dus niet bij! Wel was de kwaliteit van de karakters uitstekend, omdat voor ieder karakter een matrix van 9 bij 14 pixels ter beschikking stond. Daarnaast kon de MDA alleen monochromatisch werken, er konden dus geen kleuren worden gegenereerd.

Eigenschappen
De MDA had een eigen geheugen van 4 kB nodig voor het bewaren van de gegevens van een scherm. Dit geheugen werd geleend van het systeemgeheugen en wel op segment $B000_{hex}$. Een eigen BIOS was niet aanwezig, de routines voor het aansturen van de beeldscherm elektronica waren vastgelegd in het hoofd-BIOS van de PC zélf. Als videocontroller werd gebruik gemaakt van de 6845, die via de poortadressen $3B0_{hex}$ tot en met $3BF_{hex}$ werd aangesproken.
De resolutie bedroeg 350 lijnen van ieder 720 pixels. Er werden 50 beelden per seconde geschreven, waaruit men kan berekenen dat de horizontale sync-frequentie gelijk was aan 18.432 kHz.
Om een en ander zonder vervorming op het beeldscherm te krijgen moest de video-versterker van de monitor een bandbreedte van 16,257 MHz hebben. Vergeleken met de standaard bandbreedte van 5 MHz van een TV was dat dus al een hele uitdaging voor de monitor-ontwerpers!

Hercules

Inleiding
Reeds snel na het op de markt verschijnen van de PC met zijn MDA voelden voornamelijk professionele gebruikers de beperkingen van MDA aan den lijve. Software waarmee men kon tekenen was niet moeilijk te programmeren. Alleen had men er niets aan, want de MDA bood geen mogelijkheden om iets anders dan karakters naar de monitor te sturen. De Amerikaanse fabrikant Hercules is in dit door IBM open gelaten gat gesprongen en heeft een grafische kaart ontwikkeld, die historie heeft gemaakt: de Hercules-kaart, die een standaard geworden op het monochrome graphics gebied.
Omdat de Hercules-kaart een uitstekende tekstkwaliteit bezat en de mogelijkheid om echte "graphics" op het beeldscherm te toveren, werd deze kaart al snel een standaard in de monochrome videokaarten.
Deze standaard werd dan ook prompt, zoals gebruikelijk in de PC-wereld, door diverse hardware-leveranciers gecloond.

Eigenschappen van de Hercules-kaart
Wat maakte de Hercules kaart zo bijzonder?
Om het antwoord op deze vraag wat te illustreren volgt een korte opsomming van de belangrijkste eigenschappen van de Hercules kaart:
- de resolutie is 720 x 348 pixels;
- uitstekende letterkwaliteit door een matrix van 9 bij 14;
- het aantal tekstregels was 25 met op iedere regel ruimte voor 80 karakters;
- een video geheugen 64 kB;
- standaard parallelle printerpoort aanwezig;

- standaard TTL-monitor connector, de monitor werd dus digitaal aangestuurd met signalen van 5 V en 0 V;
- mogelijkheid zowel tekst als grafische beelden op het scherm te brengen;
- lage prijs.

Uit bovenstaande opsomming blijkt dat de Hercules kaart nogal wat voordelen had ten opzichte van andere (monochrome) video-kaarten.

De resolutie is bijvoorbeeld beter dan bij CGA-kaarten met een oplossend vermogen van slechts 640 x 200 beeldpunten in tekststand of 320 x 200 pixels in de grafische stand.

Zelfs de EGA-standaard (640 x 350 beeldpunten) moet het qua resolutie afleggen tegen een Hercules-kaart.

Videogeheugen

Van het 128 kB grote geheugensegment dat de IBM PC vrijmaakt voor de videobesturing gebruikt de Hercules-kaart als vrij videogeheugen 64 kB. Dit vrije geheugen is verdeeld in 2 zogenaamde pagina's van elk 32 kB groot. Door deze pagina-indeling van het videogeheugen is het bijvoorbeeld mogelijk om een beeld in de achtergrond op te bouwen terwijl een ander beeld nog op het scherm staat. Door nu de pagina's te wisselen kan men heel snel beelden op het scherm laten afwisselen.

Zoals al eerder vermeld beschikt de Hercules kaart over 64 kB *vrij* videogeheugen. Dit wil zeggen dat dit geheugen door de gebruiker beschreven kan worden. Deze videobuffer is verdeeld in twee zogenaamde pagina's (page 0 en 1) van elk 32 kB groot.

Stel dat de eerste pagina (page 0) actief is, dan wordt alle informatie die in dit geheugengedeelte staat, op het beeldscherm getoond. In feite komt het hier op neer, dat ieder bit in deze pagina correspondeert met een pixel op het beeldscherm. Is het bit een 1 dan verschijnt er een puntje op het beeldscherm, is het bit een 0 dan blijft het scherm op een bepaalde plaats leeg. Deze plaats van een pixel is afhankelijk van het geheugenadres van het bit. In een zeer vereenvoudigde vorm ziet zo'n beeldschermopbouw er uit zoals geschetst in figuur 3/19.9-6.

In werkelijkheid zijn de geheugenadressen van de Hercules-kaart als volgt belegd:
- page 0: B0000-B7FFF hexadecimaal;
- page 1: B8000-BFFFF hexadecimaal.

Een toepassing hiervoor kan zijn het snel wisselen van menu's. Door de ene pagina in de achtergrond op te laten bouwen terwijl de andere pagina wordt getoond kan men de bovengenoemde toepassingen realiseren.

Interface naar de video hardware

Het communiceren op machinetaalniveau met het videogedeelte van de Hercules-kaart wordt geregeld via de volgende registers:
- 03B4:
  6845 Index register;
- 03B5:
  6845 Data register;
- 08B8:
  Display Mode Control Port;
- 03BA:
  Display Status Port.

De 6845 index en data registers controleren de frequenties van de monitor en de breedte en hoogte van het scherm. Deze controle wordt met behulp van 16 parameters gedaan.

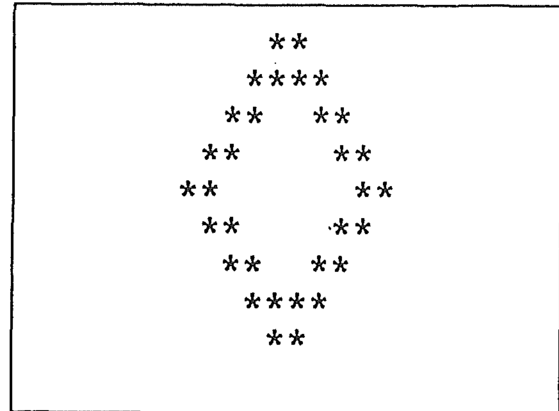
De Display Mode Control Port is het register dat de grafische operaties op de kaart laat uitvoeren. De betekenis van de 8 bits van dit register wordt samengevat in figuur 3/19.9-7.

19.9 Principes van de beeldscherm elektronica

Page 0

adres	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	1	1	0	0	0	0
1	0	0	0	1	1	1	1	0	0	0
2	0	0	1	1	0	0	1	1	0	0
3	0	1	1	0	0	0	0	1	1	0
4	1	1	0	0	0	0	0	0	1	1
5	0	1	1	0	0	0	0	1	1	0
6	0	0	1	1	0	0	1	1	0	0
7	0	0	0	1	1	1	1	0	0	0
8	0	0	0	0	1	1	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0

Beeldscherm



Figuur 3/19.9-6: De beeldschermopbouw bij de Hercules standaard.

BIT	OPTIE
0	wordt niet gebruikt
1	0 = tekstmode (bij opstarten de default) 1 = grafische mode **let op:** de 6845 registers moeten iedere keer herprogrammeerd worden wanneer dit bit van waarde verandert, doet men dit niet dan kan de monitor beschadigd worden!
2	wordt niet gebruikt
3	0 = wis het scherm (bij opstarten default) 1 = activeer het scherm gebruik dit bit bij het wisselen van mode zodat het scherm pas wordt geactiveerd wanneer de mode ingesteld is, dit zorgt voor een rustig beeldscherm
4	wordt niet gebruikt
5	0 = tekstknipperen uit (bij opstarten default) 1 = tekstknipperen aan (geen effect op cursor)
6	wordt niet gebruikt
7	0 = page 0 (bij opstarten default, start het scherm met adres B0000) 1 = page 1 (start het scherm met adres B80000)

Figuur 3/19.9-7: De betekenis van de bits van de Display Mode Control Port.

19.9 Principes van de beeldscherm elektronica

BIT	OPTIE
0	0 = normaal karakter 1 = SYNC (scherm is tijdelijk leeg)
1	wordt niet gebruikt
2	wordt niet gebruikt
3	0 = pixels uit 1 = pixels aan
4	wordt niet gebruikt
5	wordt niet gebruikt
6	wordt niet gebruikt
7	0 = vertical retrace (scherm is tijdelijk leeg) 1 = scherm is actief

Figuur 3/19.9-8: De betekenis van de bits van de Display Status Port.

De Display Status Port is het register dat als sensor werkt voor de actuele status van de monitor. De bits zijn gedefinieerd zoals aangegeven in figuur 3/19.9-8.

Interface naar de printerpoort

De communicatie tussen de Hercules-kaart en de printerpoort werd door middel van 3 registers gedaan, te weten:

- 03BC:
  Printer Data Port;
- 03BD:
  Printer Status Port;
- 03BE:
  Printer Control Port.

Aangezien de beschrijving van deze printerregisters buiten de strekking van dit hoofdstuk valt, wordt in de figuren 3/19.9-9 en -10 alleen een overzicht gegeven van de betekenis van de bits.

BIT	OPTIE
0	wordt niet gebruikt
1	wordt niet gebruikt
2	wordt niet gebruikt
3	0 = printerfout 1 = printer operatie normaal verlopen
4	0 = printer reageert niet op signalen 1 = printer reageert en luistert
5	0 = printer operatie normaal verlopen 1 = papier is op
6	0 = printer accepteert data 1 = printer wacht op meer data
7	0 = printer is bezig, niet geselecteerd of heeft een fout gedetecteerd 1 = printer operatie normaal verlopen

Figuur 3/19.9-9: De betekenis van de bits van de Printer Status Port.

19.9 Principes van de beeldscherm elektronica

BIT	OPTIE
0	0 = strobe naar printer om data op te nemen 1 = laat strobe gaan (bij opstarten default)
1	0 = auto line feed 1 = remote line feed control (bij opstarten default)
2	0 = initialiseren van printer (bij opstarten default) 1 = set printer voor normale operatie
3	0 = deselecteer de printer 1 = selecteer de printer (bij opstarten default)
4	0 = maskeer de interrupt IRQ 7 (bij opstarten default) 1 = activeer IRQ 7 wanneer de printer klaar is voor data
5	wordt niet gebruikt
6	wordt niet gebruikt
7	wordt niet gebruikt

Figuur 3/19.9-10: De betekenis van de bits van Printer Control Port.

Genereren van de beeldpunten in de tekstmode

In de tekstmode wordt het vrije videogeheugen gebruikt om de karaktercodes en de karakter attributen (intensiteit, knipperen, onderstrepen) op te slaan. Deze opslag begint bij het adres B0000 hexadecimaal en beslaat een ruimte van 4 kB tot en met adres B0FFF hexadecimaal. Het tekstscherm is 80 karakters breed en 25 regels hoog. Alle karakters worden achter elkaar opgeslagen en nemen dus maximaal 160 bytes per regel in beslag. Voor ieder karakter wordt een byte gebruikt voor de karaktercode en een byte voor de karakter attributen.

De formule voor de karaktercode is:

160*(regel-1)+2*(kolom-1)

De formule voor de attributen is:

160*(regel-1)+2*(kolom-1)+1

Hierbij is "regel" een waarde tussen 1 en 25 en "kolom" een waarde tussen 1 en 80.

De Hercules kaart heeft een karaktergenerator die 256 karakters kan genereren en een attributendecoder die een karakter kan laten onderstrepen, inverteren, wissen, vet drukken en knipperen.

Samenvatting

In de tabel van figuur 3/19.9-11 zijn de belangrijkste gegevens van de Hercules-standaard samengevat.

	Tekstmodus	Grafische modus
videosegment	b000h	b000h
grootte van video-RAM	4 kByte	64 kByte
beeldschermpagina's	1	2
videocontroller	CRTC 6845	CRTC 6845
poortadressen van 6845	3b0h-3bfh	3b0h-3bfh
tekenmatrix	9*14	9*14
effectieve tekengrootte	7*9	7*9
resolutie (pixel)	720*350	720*348
kleuren	mono	mono
monitoraansturing	digitaal	digitaal
horizontaalfrequentie[1]	18,432 kHz	18,432 kHz
verticaalfrequentie[1]	50 Hz	50 Hz
videobandbreedte[1]	16,257 MHz	16,257 MHz
eigen BIOS	nee	nee

Figuur 3/19.9-11: Samenvatting van de belangrijkste eigenschappen van de Hercules-standaard.

(M)CGA

Inleiding

In het begin van het PC-tijdperk was de keus voor video-adapters, die in staat waren om een kleurenmonitor aan te sturen, slechts beperkt tot één standaard: de in 1982 door IBM op de markt gebrachte Color Graphics Adapter (CGA).

In de loop van de tijd is deze CGA-kaart natuurlijk ook diverse malen nagemaakt door andere video-hardware leveranciers. Daarnaast heeft men geprobeerd de CGA-standaard hier en daar te "verbeteren". Zo bracht de firma Plantronics het ColorPlus-board op de markt. Dit ColorPlus-board was een zogenaamde "enhanced"-versie van de CGA-kaart, maar is nooit erg succesvol geweest. Ook IBM heeft later zelf

enkele PC-modellen met een verbeterde versie van de CGA-kaart uitgevoerd. Dit zijn de IBM PS/2 Modellen 25 en 30 die met een MultiColor Graphics Array (MCGA) waren uitgerust.

Waarom CGA?
De CGA-kaart werd in eerste instantie ontwikkeld om de nadelen van de monochrome kaarten MDA en Hercules op te heffen. Een veel geopperde wens van de gebruikers van het eerste uur was grafische mogelijkheden en tekst tegelijkertijd in kleur op het beeldscherm. Deze wens werd in vervulling gebracht toen de kaart in 1982 op de markt kwam. De PC bezat opeens grafische mogelijkheden en nog wel in verschillende standen en kleuren! Deze standen, het zijn er overigens vier, worden modes genoemd en kenmerken zich door het feit dat iedere puntje op het beeldscherm (pixel) afzonderlijk aan en uit kan worden geschakeld. Dit is al totaal verschillend met een monochrome tekstaansturing waarbij men alleen de losse karakters afzonderlijk op het beeldscherm kan manipuleren.

Kenmerken van de CGA-standaard
De CGA-standaard kent een aantal mogelijkheden om de resolutie in te stellen. Zo is er de monochrome stand waarin de CGA-video-hardware 640 x 200 beeldpunten kan aansturen. In de grafische/kleurenstand kan de CGA-kaart 320 x 200 beeldpunten in vier kleuren aansturen. Dit laatste komt neer op het tegelijkertijd op het beeldscherm aansturen van vier kleuren die uit een palet van 16 kleuren kunnen worden gekozen.

MCGA
De in de inleiding genoemde MCGA-standaard presteerde iets meer. Deze kaart bezat naast de vereiste video-hardware nog een 16 kB grote, in statische RAM uitgevoerde, tekengenerator en een beeldschermbuffer van 64 kB. Ondanks deze extra toeters en bellen kon men ook niet meer dan 320 x 200 pixels op het beeldscherm aansturen maar dan wel in 256 kleuren.
Nam men echter genoegen met 16 kleuren uit een palet van 256 kleuren dan bracht de MCGA-kaart 640 x 400 pixels op het beeldscherm.

Moeilijk leesbaar
De reden dat de CGA-standaard snel werd opgevolgd door de EGA-standaard was vooral te wijten aan de povere resolutiekwaliteit van deze video-hardware. De CGA-kaart heeft de grootste resolutie bij monochrome tekstweergave, er kunnen dan 640 x 200 beeldschermpunten aangestuurd worden. Dit betekent dat er voor 25 regels met 80 karakters slechts een letterbox van maximaal 8 x 8 pixels beschikbaar is (vergelijk bijvoorbeeld de Hercules-standaard met een letterbox van 9 x 14 pixels). Het resultaat is een korrelige, grove en slecht leesbare letter. De Color Graphics Adapter-kaart werd dan ook snel berucht om zijn vermoeiende werking bij langdurig naar het beeldscherm kijken.

De CGA-kaart intern
De CGA-kaart kan in principe in 7 standen (modes) geschakeld worden. Elk van deze standen heeft zijn eigen karakteristieke eigenschappen, hoewel enkele modes niet veel van elkaar verschillen. Een modus bepaalt hoe het beeldscherm aangestuurd wordt. Hierbij worden drie eigenschappen van de beeldschermopbouw gemanipuleerd.
Deze eigenschappen zijn:

19.9 Principes van de beeldscherm elektronica

MODE	EIGENSCHAP	RESOLUTIE	KLEUR
0	tekstweergave	80 tekens bij 25 regels	2 (zwart/wit)
1	tekstweergave	80 tekens bij 25 regels	16
2	tekstweergave	80 tekens bij 25 regels	2 (zwart/wit)
3	tekstweergave	80 tekens bij 25 regels	16
4	grafiek	320 x 200 pixels	4
5	grafiek	320 x 200 pixels	4 (grijstinten)
6	grafiek	640 x 200 pixels	2 (zwart/wit)

Figuur 3/19.9-12: De eigenschappen van de zes modes van de CGA-kaart.

- de mogelijkheid om tekst, grafiek of beide op het beeldscherm te tonen;
- het aantal tekens of pixels dat per regel op het beeldscherm kan worden geplaatst;
- de hoeveelheid kleuren die tegelijkertijd op het beeldscherm getoond kan worden.

In de tabel van figuur 3/19.9-12 worden de verschillende modes en hun eigenschappen kort samengevat.
Zoals uit deze tabel blijkt zijn er drie monochrome modes waarin de CGA-kaart kan worden geschakeld. De reden waarom dit zo geregeld is, heeft te maken met het feit dat kleuren beeldschermen vroeger zeer prijzig waren. Om de CGA-kaart toch interessant te maken voor zowel gebruikers met een monochroom als met een kleuren beeldscherm werd de CGA-kaart uitgevoerd met twee uitgangen. Een uitgang waarmee een kleurensignaal naar een kleurenmonitor kan worden verzonden en een uitgang waarmee met behulp van een kleurloos signaal een monochrome monitor kan worden aangestuurd.

Selecteren van de verschillende modes

De verschillende modes kunnen met behulp van machinetaal en BIOS aanroepen geselecteerd worden, maar de meeste hogere programmeertalen hebben geïntegreerde functies om de beeldschermmodes in te kunnen schakelen. Met behulp van deze functies kunnen de verschillende modes op eenvoudige wijze worden ingeschakeld.
Om de actuele modus te kunnen achterhalen zijn er echter nauwelijks standaard functies in de hogere programmeertalen aanwezig. Natuurlijk kan men met behulp van een BIOS-routine, namelijk de routine "Actuele Modus Opvragen", deze modus achterhalen. Er is echter een snellere methode beschikbaar om dit probleem op te lossen! Het BIOS slaat namelijk het nummer van de gebruikte modus op in een vast geheugen adres: 49_{hex}.

Het CGA beeldschermgeheugen

Uit de hierboven beschreven werkwijzen blijkt dat de CGA-kaart beschikt over een eigen geheugen dat los van het interne hoofdgeheugen wordt geadministreerd. Het hoofdgeheugen ligt tussen de adressen 0000:0000 en 9000:FFFF hexadecimaal, maar het beeldschermgeheugen van de CGA-kaart ligt tussen de adressen B800:0000 en B800:3FFF. Dit beeldschermgeheugen is dus 16 kB groot. Door in tekstmode de ASCII-codes naar het beeldschermgeheugen te schrijven worden deze tekens door de CGA-kaart op het beeldscherm getoond. Dit gebeurt doordat de CGA-kaart de ASCII-informatie vertaald naar een verzameling geheugen-

19.9 Principes van de beeldscherm elektronica

adressen die het betreffende teken dan op het beeldscherm in zijn samenstellende pixels toont. In de grafische mode wordt de pixel-informatie rechtstreeks naar het schermgeheugen geschreven en zo door de CGA-logica op het beeldscherm getoond.
Wanneer men tekst op een CGA-aangestuurde monitor wil brengen moet men in principe twee bytes manipuleren om een teken op het beeldscherm te tonen. Deze twee bytes corresponderen namelijk met een bepaald geheugenadres in het video-geheugen en dit adres correspondeert op zijn beurt weer met een karakterpositie in een regel op het beeldscherm. Het eerste byte bevat de karakter informatie terwijl het tweede byte de attribuut-informatie van het karakter bevat. Deze attribuut-informatie bestaat uit vier bits die de kleur van het teken bepalen, drie bits die de achtergrondkleur instellen en een bit dat aangeeft of het karakter helder moet worden getoond of niet.
De opbouw van het beeldschermgeheugen is in alle modes hetzelfde geregeld. Dit geheugen begint bij adres B800:0000. De ASCII-code van het teken dat in de linker bovenhoek van het scherm wordt getoond staat in adres B800:0000, de attributen van dit karakter staan in adres B800:0001. Het tweede ASCII-karakter volgt in adres B800:0002 en zijn attributen in B800:0003, enzovoort. Omdat de regellengte in de verschillende modes 40 en 80 karakters groot kan zijn is het startadres van een regel afhankelijk van de mode die men heeft gekozen. Hieruit volgt dat een mode met 80 tekens per regel (modus 2 of 3) 160 bytes beslaat (80 bytes ASCII-karakters en 80 bytes attributen) en een mode met 40 tekens per regel 80 bytes. Om nu de verschuiving van een teken in een regel ten opzichte van het eerste teken te bepalen kan men de volgende formules gebruiken:

adres = (regel*80)+(kolom*2);
modus 0, 1
of:
adres = (regel*160)+(kolom*2);
modus 2, 3

Het adres van het bijbehorende attribuut is dan **adres+1.**

Beeldschermpagina's

Ook de CGA-kaart kent net als bijvoorbeeld de Hercules-kaart de mogelijkheid om meerdere beeldschermpagina's te gebruiken. Van deze beeldschermpagina's is dan één pagina zichtbaar op het scherm, terwijl de andere beeldschermpagina's in het achtergrond geheugen kunnen worden gemanipuleerd. Hoeveel beeldschermpagina's men ter beschikking heeft, hangt wederom af van de modus die men heeft gekozen. De modes 0 en 1 gebruiken bijvoorbeeld slechts 2 kB per beeldschermpagina en kunnen dus gebruik maken van acht beeldschermpagina's die zijn genummerd van 0 tot en met 7. De modes 2 en 3 kennen 4 beeldschermpagina's omdat een beeldschermpagina nu 4 kB beslaat. De actieve pagina, dat is de pagina die op het beeldscherm wordt getoond, kan worden geselecteerd door middel van BIOS functie 5.
Natuurlijk zijn deze functies ook in hogere programmeertalen geïmplementeerd. Het manipuleren van het niet-actieve beeldschermgeheugen moet in de achtergrond gebeuren op het moment dat de actieve pagina op het beeldscherm wordt getoond. Hierdoor kon men een voor die tijd zeer snelle beeldschermwisseling creëren. Het opbouwen van het beeldschermgeheugen gebeurt door middel van het rechtstreeks schrijven naar het beeldschermgeheugen.

19.9 Principes van de beeldscherm elektronica

KLEUR	NR.	KLEURCOMPONENT ROOD	GROEN	BLAUW	INTENSITEIT (DONKER=0, HELDER=1)
zwart	0	0	0	0	0
blauw	1	0	0	1	0
groen	2	0	1	0	0
turkoise	3	0	1	1	0
rood	4	1	0	0	0
violet	5	1	0	1	0
bruin	6	1	1	0	0
grijs	7	1	1	1	0
donkergrijs	8	0	0	0	1
lichtblauw	9	0	0	1	1
lichtgroen	10	0	1	0	1
lichtturkoise	11	0	1	1	1
lichtrood	12	1	0	0	1
lila	13	1	0	1	1
geel	14	1	1	0	1
wit	15	1	1	1	1

Figuur 3/19.9-13: De samenstelling van de mengkleuren uit de basiskleuren rood, groen en blauw en de intensiteit.

De mogelijkheid om rechtstreeks naar het geheugen (en dus ook het beeldschermgeheugen) te schrijven is in iedere hogere programmeertaal geïmplementeerd.

De C van Color

Met de CGA-kaart is het in de kleurenmodes mogelijk om 16 kleuren tegelijkertijd op het beeldscherm te tonen. Deze kleuren worden uit de drie hoofdkleuren rood, groen en blauw opgebouwd. Daarnaast kan ook de intensiteit van het betreffende beeldpuntje de kleur bepalen. Is de intensiteit ingeschakeld dan verschijnt het beeldpuntje helder op het scherm, is de intensiteit niet ingeschakeld dan is het betreffende puntje donkerder. De overige kleuren turkoise tot en met wit worden zo opgebouwd door een samenstelling van deze kleuren en hun intensiteit. In de tabel van figuur 3/19.9-13 staan de 16 kleuren en hun samenstelling naast elkaar. Hoe de kleuren er nu in werkelijkheid uitzien, hangt af van de kwaliteit van de gebruikte kleurenmonitor en de persoonlijke ervaring van de gebruiker zelf.

De BIOS routines voor de CGA-aansturing

De CGA-kaart kan het eenvoudigst met behulp van de in ROM geïntegreerde Basic Input Output System (BIOS) routines aangestuurd worden.
Voor deze aansturing is de video-interrupt 10H (hexadecimaal) in het BIOS aanwezig. Door middel van het laden van een BIOS-routinenummer in het AH-register van de processor kan het beeldscherm, na aanroep van de interrupt 10H, gemanipuleerd worden. Hoewel deze routines met behulp van machinetaal programmering gebruikt kunnen worden, zijn veel hogere programmeertalen al uitgerust met routines voor beeldscherm aansturing. Voorbeelden zijn Turbo Pascal, Turbo C en GW-Basic.

19.9 Principes van de beeldscherm elektronica

De grafische mogelijkheden van de CGA-kaart

De grafische mode van de CGA-kaart is verschillend van de tekstmode omdat de beeldschermopbouw door de CGA-kaart in de grafische stand op een andere manier wordt geregeld. Het beeldschermgeheugen is anders opgedeeld. De eerste beeldschermregel staat wel in de eerste 80 bytes van het geheugen maar het adres van de tweede beeldschermregel is daar niet direct achter te vinden. De tweede regel staat namelijk in de tweede helft van het beeldschermgeheugen vanaf adres B800:2000. Na de eerste beeldschermregel staat de derde regel in de eerste helft van het geheugen en na de tweede beeldschermregel vindt men de vierde beeldschermregel in de tweede helft van het geheugen. Alle 200 regels zijn op deze manier over de twee geheugenhelften verdeeld. In alle grafische modes is deze geheugenorganisatie hetzelfde geregeld en bestaat iedere beeldschermregel uit 80 bytes. Er is door de ontwerpers van de CGA-kaart gekozen voor deze geheugenorganisatie om de video-hardware eenvoudiger te kunnen uitvoeren.

Net als voor de tekstmodes kan een formule worden ontwikkeld om het beginadres van een regel te bepalen:

adres=((regel MOD 2)*8192)+ (INT(regel/2)*80)

De vierkleuren modus 4 van de CGA-kaart was de meest gebruikte mode bij PC's met een CGA-kaart. Deze mode biedt een relatief hoge resolutie in meerdere kleuren. Elk pixel wordt in deze mode op het beeldscherm afgebeeld met behulp van twee bits. De waarden van deze bits bepalen de kleur van het pixel. Hieruit volgt direct dat er in deze mode vier kleuren mogelijk zijn. Er worden steeds vier pixels in een byte opgeslagen, zo wordt de beeldschermregel steeds met vier pixels per byte van links naar rechts opgebouwd. Met deze kennis kan men nu de routines samenstellen die een directe toegang tot het grafische beeldschermgeheugen geven. Alleen het berekenen van een adres is nu echter niet genoeg omdat verschillende pixels in een byte zijn opgeslagen. Om nu toch deze pixels te kunnen adresseren moet een bitpatroon of masker over zo'n byte gelegd worden om de juiste pixel-informatie te kunnen bepalen.

De formules zien er dan als volgt uit:

adres=((regel MOD 2)*8192)+ (INT(regel/2)*80)+(INT(kolom/8))
masker=192/(4^(kolom MOD 8))

Om nu een directe ingang tot het geheugen te krijgen kunnen bovengenoemde formules in een programma-routine worden verwerkt.

PALET	KLEUR0	KLEUR1	KLEUR2	KLEUR3
0	-	groen	rood	bruin
1	-	turkois	magenta	grijs
2	-	helgroen	helrood	geel
3	-	helturkois	lila	wit

Figuur 3/19.9-14: De samenstelling van de vier paletten van de CGA-standaard.

Het begrip palet

Om met de CGA-kaart de kleurenmogelijkheid te kunnen gebruiken is enige kennis nodig over het kleurenpalet concept. Een kleurenpalet biedt een keuze tussen alle beschikbare kleuren (CGA: 16, MCGA: 256). Bij oudere uitvoeringen van de CGA-kaart had de gebruiker de mogelijkheid om tussen twee paletten te kiezen, palet 0 en 1. In de nieuwere CGA-kaarten kon men kiezen uit 4 paletten 0 tot en met 3 waarbij de laatste twee paletten de "heldere" uitvoeringen waren van palet 0 en

19.9 Principes van de beeldscherm elektronica

1. In de tabel van figuur 3/19.9-14 zijn de paletten en hun samenstelling opgesomd. In dit overzicht is kleur0 altijd zwart maar dat kan veranderd worden naar een andere kleur. Hieruit blijkt dat in ieder palet drie kleuren vast liggen en één kleur, namelijk de achtergrondkleur, vrij gekozen kan worden.

	Tekstmodus	Grafische modus
videosegment	b800h	b800h
grootte van video-RAM	16 kbyte	16 kbyte
beeldschermpagina's	4 tot 8	1
videocontroller	CRTC 6845	CRTC 6845
poortadressen van 6845	3d0h-3dfh	3d0h-3dfh
tekenmatrix	8*8	8*8
effectieve tekengrootte	7*7, 5*7	7*7, 5*7
max. resolutie (pixel)	640*200	640*200
kleuren (max.)	16	4
monitoraansturing	digitaal	digitaal
horizontaalfrequentie[1)]	15,75 kHz	15,75 kHz
verticaalfrequentie[1)]	60 Hz	60 Hz
videobandbreedte[1)]	14,30 MHz	14,30 MHz
eigen BIOS nee	nee	

Figuur 3/19.9-15: De belangrijkste kenmerken van de CGA-standaard.

High resolution modus 6
De belangrijkste eigenschap van de Hires-mode is dat elke regel nu uit 640 pixels is opgebouwd. Ook hier wordt elk pixel als een bit opgeslagen zodat het gebruik van het beeldschermgeheugen niet verschilt van de andere modes. Ook voor modus 6 zijn er formules te ontwikkelen om een beeldscherm geheugenadres te bepalen. De formule voor het byte adres is identiek aan de formule van modes 4 en 5:

adres=((regel MOD 2)*8192)+ (INT(regel/2)*80)+(INT(kolom/8))

Alleen de formule voor het masker is veranderd omdat nu het masker slechts één bit hoeft te beslaan:

masker=Shr($80,(kolom MOD 8))

Om het beeldschermgeheugen te manipuleren kan men op dezelfde manier te werk gaan als bij de modes 4 en 5. Hoewel een pixel in modus 6 in principe slechts twee kleuren kan hebben, namelijk zwart (pixel is uit) en wit (pixel is aan) was het met de laatste generatie CGA-kaarten mogelijk om de kleur van de pixels op het beeldscherm te bepalen.

Samenvatting
De belangrijkste kenmerken van de CGA-standaard zijn samengevat in de tabel van figuur 3/19.9-15.

EGA en PGC

Voorgeschiedenis
In de evolutie van video-kaarten voor de PC, verschijnt in 1984 de Enhanced Graphics Adapter (EGA) op het toneel. Deze nieuwe videokaart werd in dat jaar door IBM in de PC-AT geïntroduceerd om in de groeiende behoefte aan goede grafische voorzieningen te voldoen. IBM had in het verleden al een steek laten vallen door de monochrome MDA-standaard niet van een grafische mode te voorzien. Hierdoor kon een bedrijf als Hercules snel inspelen op deze misser van IBM met een monochrome videokaart die wel grafiek op het beeldscherm kon weergeven. Ook de CGA-kaart liet qua prestaties veel te wensen over, zodat IBM met deze EGA-standaard veel goed moest maken.

Nieuwe mogelijkheden
Naast de bekende monochrome tekst weergave en de CGA-modes bood de EGA-standaard een viertal nieuwe mogelijkheden. De meest bekende en gebruikte hiervan is de eigenschap om grafiek met een resolutie van 640 x 350 beeldpunten op het scherm te brengen en deze

beeldpunten ook nog eens in 16 verschillende kleuren tegelijkertijd weer te geven. Hoewel dit laatste veelbelovend was, kleefden er aan de EGA-videokaart wel een paar stevige nadelen.
Niet alleen was de kaart bij zijn introductie erg duur, maar ook de alom vertrouwde RGB-monitor kon niet meer in combinatie met een EGA-videokaart gebruikt worden. De EGA-videokaart geeft namelijk een hogere frequentie af waardoor een andere monitor moet worden gebruikt. Tegelijk met de introductie van de EGA-videokaart bracht IBM dan ook een Enhanced Color Display op de markt. Deze monitoren (en gelijksoortige monitoren van andere leveranciers) waren echter ook zeer prijzig.
Een derde nadeel van de EGA-videokaart was dat de adapter pas naar behoren functioneerde in een PC-AT op basis van een Intel 80286 processor. De verwerkingssnelheid van een standaard PC op basis van een Intel 8086/8088 processor was eenvoudig niet hoog genoeg om de beeldschermopbouw van de EGA-kaart snel af te kunnen handelen.
Door deze nadelen leek het erop dat de EGA-kaart in de periode na zijn introductie geen lang leven beschoren zou zijn. Maar door forse prijsdalingen van zowel EGA-kaarten als EGA-monitoren (en niet te vergeten de PC-AT's) kwam hier al snel verandering in.

MultiSync-monitoren
Een andere ontwikkeling die zeker aan het succes van de EGA-videokaart heeft bijgedragen, is de komst van de NEC MultiSync monitor geweest. Deze monitor kan een hele range van verschillende (hogere) frequenties verwerken en schakelt automatisch over naar de door de videokaart afgegeven frequentie.

Wat was de EGA-standaard?
Een beeld op een EGA-monitor zag er, vooral in vergelijking met een Hercules of CGA-monitor, prachtig uit. Zestien heldere kleuren, te kiezen uit een palet van 64 kleuren. En dit allemaal tegelijk op het scherm met een prima resolutie! Ongekende luxe voor die tijd! Vooral programma's die veel gebruik maken van kleur en grafische schermen gingen er zichtbaar op vooruit wanneer ze in combinatie met een EGA-kaart werden gebruikt. Men kan dan ook zonder meer stellen dat de introductie van de EGA-standaard een flink stuk heeft bijgedragen aan de doorbraak van de PC als grafische machine. Denk daarbij bijvoorbeeld maar eens aan CAD- en tekenprogramma's en programma's die gebruikt worden voor het maken van animaties, presentaties en grafieken. Een EGA-kaart deed daarbij geen concessies aan de tekstweergave: de resolutie is met 640 x 350 beeldpunten vrijwel vergelijkbaar met de resolutie van een monochrome tekstbeeldscherm met 720 x 350 beeldpunten. De tekstbox van de EGA-videokaart kan naast 8 x 8 beeldpunten ook uit 8 x 14 pixels bestaan.
Verder kan de EGA-videokaart naast de gebruikelijke 25 regels ook 43 regels van 80 karakters op het beeldscherm tonen.

De originele IBM EGA-videokaart
De oorspronkelijke IBM EGA-videokaart was uitgerust met 64 kB video RAM-geheugen op de kaart. Dit was echter niet genoeg om het beeldscherm volledig aan te sturen. Om toch de benodigde 256 kB te kunnen gebruiken werden 192 kB van het systeemgeheugen genomen. Latere EGA-kaarten van andere leveranciers werden geconfigureerd met de benodigde 256 kB standaard op de kaart geïntegreerd. De EGA-videokaart bezit boven-

dien nog 16 kB aan ROM-geheugen ter vervanging en aanvulling van de videoroutines van het BIOS.

Compatibiliteit
De EGA-videokaart bezat naast de specifieke EGA-modes ook de eigenschappen van de MDA- en CGA-videokaart. Dit is gedaan om de compatibiliteit met oudere software te waarborgen.

De EGA-kaart en zijn clonen
Zoals hierboven al vermeld leverde IBM zijn EGA-kaart met slechts 64 kB aan videogeheugen op de kaart. De rest van het benodigde geheugen werd van het systeemgeheugen afgesnoept. Om dit manco te verhelpen kwamen er al snel EGA-kaarten van andere leveranciers op de markt met de volledige 256 kB aan video-RAM op de kaart geïnstalleerd. Naast kaarten met het extra geheugen kwamen er ook steeds meer EGA-clonen die met sterk geïntegreerde hardware-componenten uitgerust waren. Was de video-logica bij de oorspronkelijke EGA-kaart nog verdeeld over diverse chip's, bij de modernere EGA-clonen werden de meeste hoofdfuncties van de EGA-logica in één chip ondergebracht. Ook dit was een belangrijke doorbraak, want die integratie van steeds meer video-functionaliteit in steeds minder chip's is een proces dat tot op de dag van heden doorgaat. Veel bekende chip-fabrikanten, zoals Intel met de 82786 videochip, Texas Instruments met zijn TI34010 of Hitachi met het 63484 Viking 1-board, presteerden het de volledige EGA-logica in één chip onder te brengen (de zogenaamde single-chip solutions).

Professional Graphics Controller
Tegelijkertijd met de standaard EGA-kaart, bracht IBM ook de zogenaamde "Professional Graphics Controller" (PGC) op de markt. Deze "video-adapter voor de CAD-werker" bood een resolutie van 640 x 480 pixels. Het logisch gevolg was dat er EGA-clonen op de markt kwamen, de zogenoemde "EGA-Plus" kaarten, die zowel de oorspronkelijke EGA-resolutie van 640 x 350 pixels als ook de PGC-resolutie (640 x 480) op het scherm konden brengen.

De monitor-revolutie
Zoals reeds geschreven werkte de EGA-kaart niet met de standaard TTL- of RGB-monitoren die gebruikelijk waren. De EGA-kaart gebruikt namelijk hogere scanfrequenties om een monitor aan te sturen. Zo is de horizontale aftastfrequentie (horizontal scan frequency) van een standaard kleur/grafiek-kaart 15,75 MHz, terwijl deze voor een EGA-kaart 21,85 MHz bedraagt. Een gewone kleurenbuis kan die 21,85 MHz niet bijbenen.
Een monitor die zeker bijgedragen heeft aan het uiteindelijke succes van de EGA-kaart, was ongetwijfeld de MultiSync monitor van de firma NEC. Deze monitor kostte bij zijn introductie bijna drieduizend gulden. Voor dit geld kreeg men echter een monitor die bruikbaar was in combinatie met een veelheid aan videosystemen, zowel monochrome als kleur. NEC ontwikkelde met de MultiSync een revolutionair nieuw concept. Een monitor die zich automatisch instelt op de door de video-elektronica van een computer afgegeven scanfrequentie. De originele NEC MultiSync kon zelfs nog veel hogere frequenties aan dan nodig voor de EGA-standaard. Met een resoluties van 800 x 600 pixels had MultiSync geen moeite. De MultiSync heeft model gestaan voor alle huidige monitoren van alle merken. Moderne MultiSync-monitoren kunnen met

gemak een resolutie van 1.600 bij 1.200 aan, waarbij vertikale beeldverversingfrequenties van 90 Hz eerder regel dan uitzondering zijn!
Samengevat staan de belangrijkste kenmerken van de geavanceerde EGA-monitoren hier nog eens onder elkaar:
- automatisch bijstellen van de horizontale synchronisatie aan de hand van de gebruikte video-adapter-kaart;
- een naar keuze digitale of analoge display-mode;
- variabele scanfrequenties van 14 à 16 MHz tot 20 à 26 MHz.

De EGA video-hardware
In tegenstelling tot de CGA- en Hercules-standaarden maakt de EGA video-hardware geen gebruik van de Motorola 6845 CRTC-chip (CRTC = Cathode Ray Tube Controller). De EGA maakt echter wel gebruik van een vergelijkbare chip die gebaseerd is op het ontwerp van deze 6845-chip. De voor de EGA gebruikte chip is speciaal op maat gemaakt voor zijn EGA-taak. Een gevolg van deze aanpak is dat de EGA niet geheel compatibel was met videosystemen die wel de oorspronkelijke 6845-chip gebruikten.
Een tweede verschil tussen de EGA-videohardware en bijvoorbeeld de CGA- of Hercules-standaarden, is dat verschillende functies die bij de laatste twee videosystemen in één chip geïntegreerd zijn, bij de (oorspronkelijke) EGA-standaard over meerdere chip's zijn verdeeld. Daarnaast bezit de EGA CRTC-chip zes extra registers om het beeldscherm te kunnen manipuleren.
De EGA kan uitgerust zijn met 64, 128 of 256 kB aan intern RAM-geheugen. Naast de functie als videobuffer, biedt het RAM-geheugen ook plaats aan de gegevens voor de bitpatronen van 1.024 tekens. Het beginadres van de buffer is programmeerbaar en begint voor de speciale EGA grafische modes op A0000H en ten behoeve van de compatibiliteit met de monochrome en de color-graphics standaarden op respectievelijk adres B0000H en B8000H.
De EGA-standaard neemt nooit meer adresruimte in beslag dan de twee geheugensegmenten van A000H tot BFFFH, zelfs als de kaart over 256 kB aan intern geheugen beschikt.
Dit is mogelijk omdat in enkele videomodes twee of meer bytes van het videogeheugen benaderd worden via hetzelfde geheugenadres. Het aantal beschikbare pagina's is afhankelijk van de schermmode en de hoeveelheid intern geheugen.
Verder beschikt de EGA over 16 kB ROM-geheugen ter vervanging en aanvulling van de videoroutines van het BIOS. Het ROM geheugen begint op C000:0000.

EGA-kleuren
In de grafische mode wordt naast de positie van de beeldschermen ook de kleurinformatie in de buffer opgeslagen. In de EGA-kaart bevat elk deel van de buffer één bit van de twee of vier bits die samen de kleur van een punt op het scherm bepalen. De EGA-kaart kan op een scherm maximaal 16 kleuren uit een palet van 64 mogelijke kleuren afbeelden. Als er 16 kleuren worden gebruikt, zijn voor elke pixel vier geheugenbits nodig (vier bits kunnen een getal bevatten van 0 tot en met 15). Voor grafische afbeeldingen in twee kleuren zijn echter maar twee bits per pixel nodig. In het geval van grafische afbeeldingen in twee kleuren kunnen acht pixels in een enkele byte in de videobuffer worden opgeslagen. De hoeveelheid geheugenruimte die voor een bepaald type scherm nodig is, is eenvoudig te berekenen door het totale aantal pixels

19.9 Principes van de beeldscherm elektronica

te vermenigvuldigen met de hoeveelheid bits per pixel.
Tekst en grafische afbeeldingen zijn zonder problemen op een scherm te kombineren (het BIOS tekent de tekens op het grafisch scherm) en daardoor zijn speciale tekens die niet in de standaard tekenset bestaan zelf te creëren.

De EGA besturingsregisters

Naast de 18 compatibele registers die de 6845-chip (Hercules- en CGA-standaarden) heeft, is de op de EGA-kaart gebruikte CRTC-chip uitgerust met nog zes aanvullende registers. De meeste van deze aanvullende registers regelen een aantal interne technische zaken. Registernummer 20 is het interessantst omdat dit register de scanlijn van de onderstreping van tekens regelt en zodoende de vorm van de onderstreping bepaalt.
Bovengenoemde registers worden net als bij de Hercules- en CGA-standaarden via respectievelijke de poortadressen 3B5H en 3D5H benaderd. Het verschil met de Hercules/CGA- en de EGA-standaard ligt hierin, dat bij EGA afhankelijk van de aangesloten monitor (monochroom of kleur) poortadres 3B5H of 3D5H wordt gebruikt.

EGA-modes

Van de Hercules-, CGA- en EGA-standaard heeft de EGA de meeste schermmodes. In de tabel van figuur 3/19.9-16 worden alle verschillende modes (Hercules, CGA en EGA) opgesomd.
In mode 7, de gewone monochrome tekstmode, staat de EGA het gebruik van acht pagina's toe. De modes 0 tot en met 6 zijn volledig compatibel. Zij maken op dezelfde manier gebruik van het geheugen als de andere kaarten. Met behulp van configuratieschakelaartjes, zogenaamde dipswitches, die op de EGA-kaart waren geïntegreerd kon de kaart voor een IBM Enhanced Color Display worden ingesteld. Is dit het geval dan worden de traditionele tekstmodes afgebeeld in kleur en in een hoge resolutie. De tekens bestaan dan namelijk uit 8 x 14 pixels in plaats van de gebruikelijke 8 x 8 pixels. In het bijzonder kunnen in deze uitgebreide mode de tekens in kleur onderstreept worden, de 8 x 14 pixels grote tekencel beschikt daarvoor over een extra scanlijn.

De EGA paletregisters

Om de kleuren op een EGA-beeldscherm te kunnen weergeven maakt de EGA-kaart gebruik van 16 paletregisters. Deze registers bevinden zich op poortadres 3C0H en de paletnummers lopen van 00 tot 0FH. Om toegang te krijgen tot het adresregister op 3C0H, moet men eerst poort 3DAH uitlezen. Vervolgens moet het nummer van het paletregister naar poort 3C0H worden geschreven. En uiteindelijk worden de gegevens naar hetzelfde adres geschreven.
Als de dipswitches op de EGA-kaart zijn ingesteld voor de enhanced mode, kan het palet worden samengesteld uit 64 kleuren. In dat geval zijn de instellingen voor het paletregister zes bits lang en hebben ze de vorm R'G'B'-RGB. De RGB-bits voorzien in donkere kleuren en de R'G'B'-bits leveren heldere kleuren. Als zowel bijvoorbeeld R als R' gelijk zijn aan 1, levert dat een heldere rode kleur. Door de bits te mengen worden allerlei tinten verkregen.
Als de paletregisters zijn ingesteld voor 64 kleuren terwijl de EGA niet in de uitgebreide mode staat, worden de bits 4 en 5 van de registers genegeerd en wordt de inhoud van de registers behandeld als een gewoon IRGB-patroon.

19.9 Principes van de beeldscherm elektronica

Nummer	Mode	Kaart
0	40x25 (320x200) Z/W alfanumeriek	CGA, EGA
1	40x25 (320x200) kleur alfanumeriek	CGA, EGA
2	80x25 (640x200) Z/W alfanumeriek	CGA, EGA
3	80x25 (640x200) kleur alfanumeriek	CGA, EGA
4	320x200 grafisch, 4 kleuren	CGA, EGA
5	320x200 grafisch, Z/W	CGA, EGA
6	640x200 grafisch, Z/W	CGA, EGA
7	80x25 (720x350) Z/W alfanumeriek	monochrome, Hercules en EGA
B	gereserveerd voor EGA	
C	gereserveerd voor EGA	
D	320x200 grafisch, 16 kleuren	EGA
E	640x200 grafisch, 16 kleuren	EGA
F	640x350 grafisch, 4 kleuren op monochrome monitor	EGA
10	640x350, grafisch, 4 of 16 kleuren	EGA

Figuur 3/19.9-16: De verschillende schermmodes van de EGA-standaard.

Nr	Rood	Groen	Blauw	Kleur	Nr	Felle kleur
0	0	0	0	Zwart	8	Grijs
1	0	0	1	Blauw	9	Lichtblauw
2	0	1	0	Groen	10	Lichtgroen
3	0	1	1	Cyaan	11	Lichtcyaan
4	1	0	0	Rood	12	Lichtrood
5	1	0	1	Magenta	13	Lichtmagenta
6	1	1	0	Bruin	14	Geel
7	1	1	1	Wit	15	Helder wit

Figuur 3/19.9-17: Het verband tussen de kleurnummers en de kleuren op het scherm.

Omdat de EGA gebruik maakt van paletregisters is de keuze aan achtergrondkleuren niet beperkt door het gebruik van bit 7 van een attribuutbyte als knipperbit. De codes voor de voorgrondkleur lopen van 0 tot 31. De nummers 0 tot 15 komen overeen met de kleuren in de tabel van figuur 3/19.9-17. De nummers 16 tot 31 komen overeen met dezelfde kleuren als in de tabel maar dan worden zij knipperend op het scherm weergegeven.

Het plaatsen van een teken op het scherm

Alle BIOS- en DOS-functies die tekens naar het scherm schrijven, plaatsen dat teken op de huidige cursorpositie en schuiven vervolgens de cursor een plaats op. Ook verplaatsen al deze functies de cursor van het einde van een regel naar het begin van de volgende, tenzij het nadrukkelijk zo geregeld is dat alle tekens die voorbij de laatste (tachtigste) kolom terecht komen genegeerd worden. Sommige functies besteden wel aandacht aan het attribuut van het teken, andere niet. De buffer van de monochrome kaart en de Hercules-kaart begint op B000:0000 en die van de CGA-kaart op B800:0000. De EGA maakt in overeenkomstige modes gebruik van dezelfde beginadressen. Adressen met een even nummer (met 0 als eerste) bevatten de ASCII-codes en de oneven adressen de attribuutbytes. Wanneer men rechtstreeks naar het videogeheugen schrijft wordt de cursor niet verplaatst. Uitzetten van de cursor is een optie om verwarring te voorkomen. Wel moet zelf een pointer worden bijgehouden om te kunnen bepalen waar men in het videogeheugen is gebleven.

(wordt vervolgd)

19.9 Principes van de beeldscherm elektronica

Speciale tekens ontwerpen

Met de EGA-kaart is het net als bij de CGA-kaart mogelijk om speciale tekens te ontwerpen. Zo staat de CGA-kaart 128 door de gebruiker gedefinieerde tekens toe en de EGA zelfs 1.024, waarvan er 512 tegelijk te gebruiken zijn. De ROM-BIOS van de CGA-kaart bevat alleen de gegevens voor de vorm van de eerste 128 tekens van de ASCII-set (nummers 0 tot 127). De tweede 128 zijn niet beschikbaar tenzij ze door de gebruiker zelf worden ontworpen. Het besturingssysteem DOS beschikt sinds versie 3.0 over de instructie GRAFTABL. Met deze opdracht kunnen de vereiste gegevens voor de tweede reeks van 128 tekens worden geleverd.

De EGA beschikt over de komplete tekenset voor zowel de 25-regel als de 43-regel modes. In tegenstelling tot de eenvoudige werkwijze bij de CGA-kaart om tekens te ontwerpen is het bij de EGA wat complexer om dit te programmeren. Wel heeft de EGA heel wat meer mogelijkheden om zelf tekens te ontwerpen. Als de EGA in tekstmode wordt gezet, wordt een van de twee tekensets (8 x 8 of 8 x 14) vanuit de ROM van de EGA naar bit-map 2 in de videobuffer gekopieerd. Dit deel van de buffer wordt behandeld alsof hij uit blokken bestaat en de gewone tekenset staat in blok 0. Wanneer de EGA-kaart over genoeg geheugen beschikt kunnen er nog drie van deze geheugenblokken met tekensetinformatie worden gevuld. De grootte van een blok is afhankelijk van het aantal gebruikte scanlijnen, 8 of 14. Voor de tekens die uit 8 x 8 beeldpunten bestaan, zijn 8 x 256 = 2.048 bytes nodig. Als er meer dan een blok met tekens is ingesteld, bepaalt bit 3 van het attribuutbyte welk geheugenblok in de videobuffer gebruikt wordt om de tekensetinformatie te leveren.

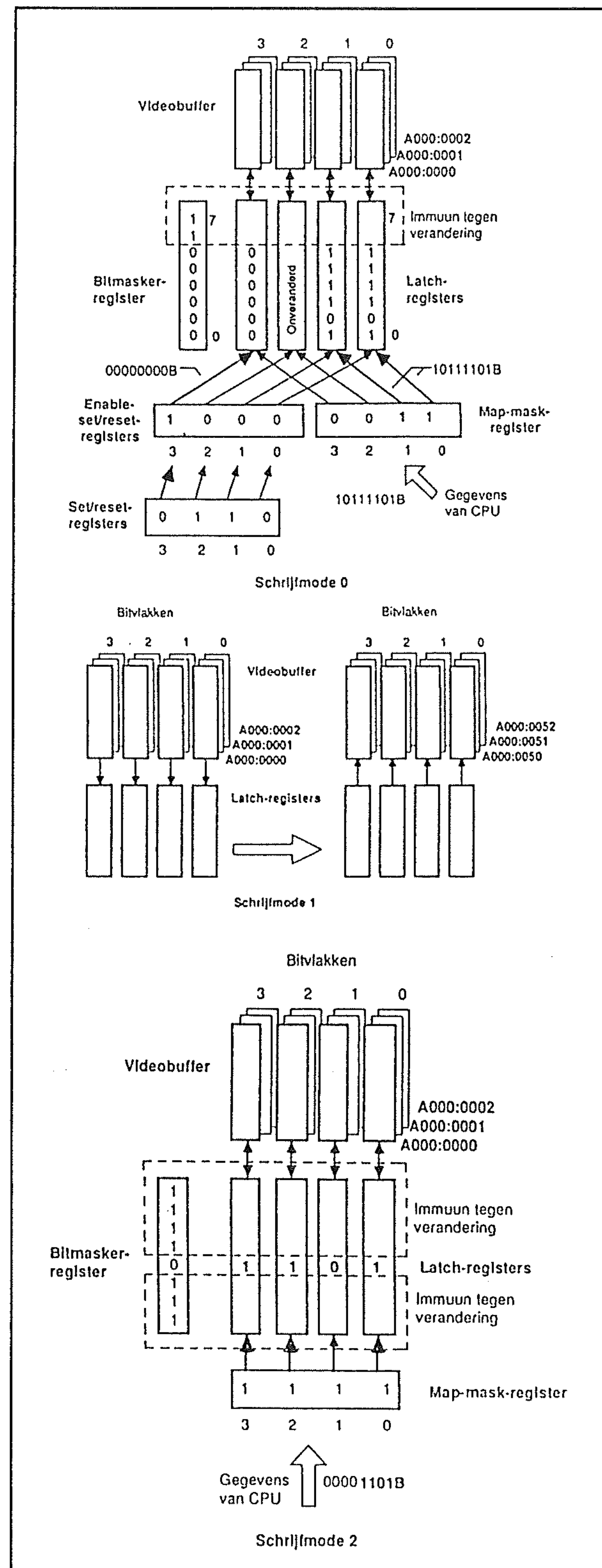
Welk geheugenblok wordt gebruikt is afhankelijk van de instelling van de bits 0 tot 3 van het character-map-select-register (tekenset keuzeregister). Dit register bevindt zich op het poortadres 3C5H. Stuur eerst 3 naar het poortadres 3C4H om het register te indexeren. Wanneer bit 3 van het attribuutbyte gelijk is aan 0 dan wordt het blok gebruikt dat wordt aangewezen door de bits 0 en 1 van het keuzeregister. Is bit 3 gelijk aan 1, dan wordt het geheugenblok gebruikt dat wordt aangewezen door bits 2 en 3. Als de bits 0 en 1 gelijk zijn aan de bits 2 en 3, dan is de mogelijkheid om twee tekenset tegelijkertijd te gebruiken uitgeschakeld en wordt bit 3 gebruikt om de helderheid van het teken aan te geven. In dat geval is alleen blok 0 in gebruik om de tekensetinformatie aan te leveren. Natuurlijk kunnen op iedere willekeurig plaats in dit blok zelf ontworpen tekens worden tussengevoegd. De oorspronkelijke tekensetinformatie wordt dan weliswaar overschreven, maar deze informatie is te allen tijde weer uit de standaard tekensetinformatie vanuit het ROM te herstellen.

Tekenen van een punt op het scherm

Naast de uit compatibiliteitsredenen aanwezige grafische CGA- en de monochrome modes heeft de EGA-kaart nog een aantal uitgebreide grafische modes (DH tot 10H). In deze modes DH, EH en 10H is het geheugen georganiseerd in de vorm van 4 bitvlakken. Een enkel bitvlak is op dezelfde manier georganiseerd als de zwart/wit-mode met hoge resolutie op de CGA. Als er een gegevensbyte naar een adres in de videobuffer wordt gestuurd, komt elk bit overeen met een pixel op het scherm en wel in de vorm van een horizontaal element met bit 7 aan de linkerkant. De bitvlakken zijn in principe zo

19.9 Principes van de beeldscherm elektronica

gerangschikt dat de vlakken boven elkaar op hetzelfde adres in de videobuffer staan. Op deze manier zijn er vier bits voor elk punt die zo 16 kleuren opleveren. Om de gegevens in de bitvlakken te kunnen benaderen bezit de EGA-elektronica een aantal registers, waaronder twee masker-registers en vier latchregisters. Deze latch-registers bevatten de gegevens van elk van de bitvlakken voor de geheugenplaats die het laatst is benaderd. Als de CPU van de computer gegevens naar een bepaald adres stuurt, kunnen die gegevens de oude gegevens in de latchregisters geheel of gedeeltelijk vervangen en vervolgens worden dan de gegevens in de latchregisters naar de videobuffer verplaatst. In de latchregisters worden de CPU-gegevens veranderd afhankelijk van de ingestelde grafische schrijfmode. Bij iedere leesactie van een videobuffer-adres worden de latchregisters geladen met de vier bytes in de vier bitvlakken op die plaats. Door nu de latchregisters te bewerken met eenvoudige instructies als OR, AND, XOR of bitrotaties, kunnen prachtige grafische en/of scroll effecten worden verkregen. De twee maskerregisters, een bit-masker-register en een map-maskerregister beschermen bepaalde bits of bitvlakken tegen veranderingen door de gegevens die van de CPU afkomstig zijn. Het bit-maskerregister bevindt zich op adres 3CFH en kan uitsluitend beschreven worden. Om dit register te kunnen beschrijven moet de waarde 8H naar poort 3CEH gestuurd worden om het bit-maskerregister te indexeren. Door nu een bit in dit register gelijk aan 1 te maken wordt het overeenkomstige bit in alle vier de bitvlakken beschermd tegen overschrijven en zo ook het betreffende pixel op het beeldscherm. Deze werkwijze wordt maskeren genoemd.



Figuur 3/19.9-18: De schrijfmodes van de EGA-kaart.

19.9 Principes van de beeldscherm elektronica

Omdat de videohardware alleen met bytes werkt en niet met bits, worden de videogegevens eerst in de latchregisters geladen.
Het map-maskerregister werkt op eenzelfde manier als het bit-maskerregister, maar wordt door de verschillende grafische schrijfmodes op uiteenlopende manieren gebruikt. De schrijfmodes van de EGA worden door het moderegister ingesteld. Dit register is ook weer alleen te beschrijven en wel op poortadres 3CFH. Daarvoor moet de waarde 5H naar de poort worden gestuurd om het register te kunnen indexeren. Van dit register worden bit 0 en 1 gebruikt om de schrijfmode in te stellen. Bit 3 van dit register wordt gebruikt om een van de twee leesmogelijkheden van de videobuffer in te stellen. De overige bits in het moderegister moeten altijd 0 zijn. De EGA-BIOS stelt standaard de schrijfmode 0 in, daarnaast zijn er de modes 1 en 2.

	Tekstmodus	Grafische modus
videosegment	b800h	a000h
grootte van video-RAM	64-256 kbyte	64-256 kbyte
beeldschermpagina's	1 tot 8	1 tot 8
videocontroller	EGA-CRTC	EGA-CRTC
poortadressen	3d0h-3dfh	3c0h-3dfh
tekenmatrix	8*14,8*8	8*14,8*8
effectieve tekengrootte	7*9,7*7	7*9,7*7
resolutie (pixel)	640*350	640*350
kleuren	16 van 64	16 van 64
monitoraansturing	digitaal	digitaal
horizontaalfrequentie	15,7-21,8 kHz	15,7-21,8 kHz
verticaalfrequentie	60 Hz	60 Hz
videobandbreedte	14,3-16,3 MHz	14,3-16,3 MHz
eigen BIOS	ja	ja

Figuur 3/19.9-19: De belangrijkste specificaties van de EGA-standaard.

De grafische schrijfmodes

De werking van de grafische schrijfmodes is samengevat in het overzicht van figuur 3/19.9-18.

Samenvatting

In de tabel van figuur 3/19.9-19 zijn de belangrijkste specificaties van de EGA-standaard samengevat.

VGA

Inleiding

In de technologische ontwikkeling van beeldscherm aansturingskaarten voor IBM-compatibele personal computers was de "Video Graphics Array", of kortweg de VGA-standaard, een nieuwe doorbraak. Het beeldscherm van de PC oogde volwassen en PC-bezitters moesten zich niet langer schamen als zij de prachtige kleuren en grafische mogelijkheden van andere platforms zoals Atari of Apple IBM wel eens onder ogen kregen. IBM introduceerde de VGA-standaard in april 1988, tegelijkertijd met de introductie van de PS/2 modellen 50, 60 en 80. Kort daarna werd óók deze nieuwe standaard door onder andere aziatische PC-producenten gekopieerd en in de bekende "klonen" ingebouwd. Het gevolg was dat binnen de kortste keren VGA-kaarten in alle soorten en maten werden aangeboden. Superlatieven als "super" en "enhanced" werden gebruikt om de vele extra's en specifieke eigenschappen van zo'n kaart aan te kunnen duiden. De oorspronkelijke VGA-standaard is echter nog steeds niet van het toneel verdwenen. Door de steeds verdergaande integratie van de verschillende componenten van bijvoorbeeld een VGA-kaart komt men de "standaard" VGA-uitvoering nog steeds tegen in goedkope laptop en notebook computers. Hierbij is dan de gehele videotechnologie in een of enkele chip's ondergebracht en op het moederbord gemonteerd.

19.9 Principes van de beeldscherm elektronica

De opvolger van de EGA-standaard

De allereerste uitvoering van de VGA-kaart was eigenlijk niet meer dan een EGA-kaart, gebakken in nieuwe, modernere chip's. De VGA-kaart kreeg echter een hogere resolutie en de oude EGA-bug's werden verholpen. Daarnaast werden zowel in high als in low resolution het maximale oplossend vermogen verhoogd. Omdat er voor de EGA-kaart inmiddels veel software beschikbaar was, kon IBM er echter niet omheen om de VGA-kaart volledig compatibel met de EGA-kaart te maken.

Analoge aansturing

Een echt groot verschil tussen VGA en EGA ligt hierin dat de VGA-technologie het beeldscherm analoog aanstuurt terwijl EGA dat nog altijd digitaal oplost. Het analoog aansturen is pure noodzaak om onder andere alle gewenste kleuren op het scherm te kunnen brengen, dit komt met name de beeldkwaliteit ten goede.

Eigenschappen

De resolutie van de standaard VGA-kaart is evenals de EGA-kaart te verdelen in verschillende modes (zie figuur 3/19.9-20). Elke mode heeft zijn eigen kenmerken wat betreft het oplossend vermogen. Voor tekst kan de VGA-kaart in drie modes worden geschakeld, namelijk:

- 720 x 400 beeldpunten-mode in 16 kleuren;
- 720 x 400 beeldpunten-mode in 2 kleuren;
- 360 x 400 beeldpunten-mode wederom in 16 kleuren.

In de eerste en laatste mode zijn de 16 kleuren te kiezen uit een palet van 256 kleuren.

In high resolution graphics (HIRES) heeft de VGA-kaart per horizontale scanlijn 640 pixels en per vertikale scanlijn 480 pixels ter beschikking. In deze grafische stand is de VGA-kaart in twee modes te schakelen. Een mode met 16 kleuren die tegelijkertijd op het scherm te gebruiken zijn en uit een palet van 256 kleuren kunnen worden gekozen. Tevens bestaat er nog een mode met twee kleuren in dezelfde resolutie.

In low resolution graphics (LOWRES) zijn slechts 320 x 200 pixels te adresseren, maar dan zijn er wel 256 kleuren tegelijkertijd te gebruiken, uit een kleurenpalet van 262.144 tinten.

Resolutie	Mode	Kleuren	HSF (kHz)	VSF (Hz)
720 x 400	Tekst	16	70	31,50
720 x 400	Tekst	2	70	31,50
360 x 400	Tekst	16	70	31,50
640 x 480	Grafiek	16	60	31,50
640 x 480	Grafiek	2	60	31,50
320 x 200	Grafiek	256	70	31,50

Figuur 3/19.9-20: De standaard VGA-resoluties en de gebruikte scanfrequenties.

Daarnaast kunnen computers, uitgerust met een monochrome monitor, zo worden geconfigureerd dat 64 grijstinten op het beeldscherm kunnen worden onderscheiden. Deze monochrome techniek werd voornamelijk toegepast in de draagbare laptop en notebook computers van de eerste generatie, die uitgerust waren met monochrome LCD-beeldschermen.

Het VGA videogeheugen

De standaard VGA-kaart was bij zijn introductie uitgerust met 256 kB aan Video-RAM (VRAM) gemonteerd op de kaart. Tegenwoordig zijn deze kaarten veelal uit-

19.9 Principes van de beeldscherm elektronica

gevoerd met 512 kB tot 2 MB aan Video-RAM op de kaart. Algemeen geldt, net als bij de EGA-standaard, dat naarmate er meer Video-RAM op de kaart aanwezig is, er minder RAM van het computergeheugen hoeft te worden gesnoept. Er worden bij een standaard VGA-beeld nog altijd 640 x 480 = 307.200 pixels opgeslagen! Dit vraagt, met 8 bits per pixel, aan opslagcapaciteit een kleine 300 kB. 256 kB aan Video-RAM schiet hier dus duidelijk in tekort. Het grote voordeel van Video-RAM ten opzichte van gewoon DRAM-geheugen zit in de snelheidswinst. Video-RAM is sneller dan DRAM, maar hierover later meer.

De technische eigenschappen
Omdat voor de standaard VGA-kaart de technische eigenschappen aan de ene kant vrijwel zijn uitgekristalliseerd en aan de andere kant van kaart tot kaart verschillen, kunnen deze kaarten het best onderling worden vergeleken aan de hand van de snelheidsspecificaties van de kaart.
De snelheidsspecificaties worden door een aantal factoren beïnvloed. Hierbij spelen de kaartconnector, het videogeheugen, aanwezigheid van een videocoprocessor en de kwaliteit van de software-driver een belangrijke rol.

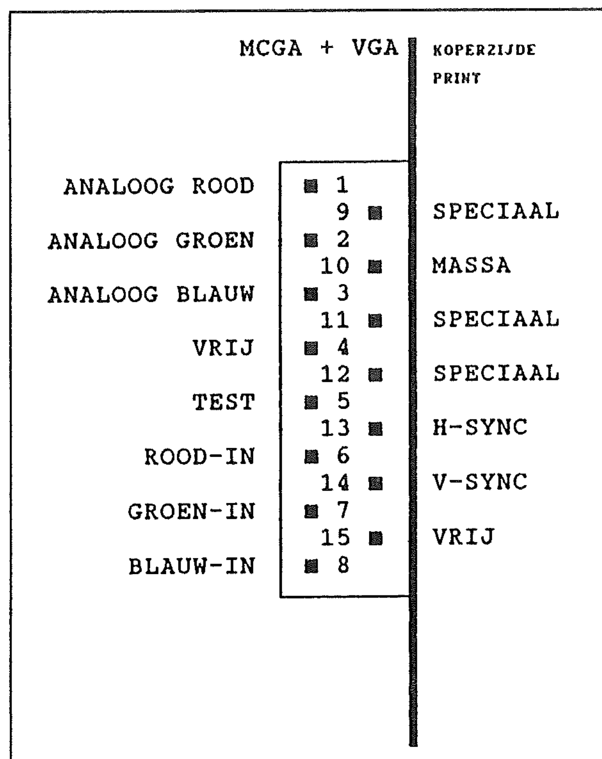
De kaartconnector en de busfactor
De kaartconnector kan 8 of 16 bit groot zijn, dit hangt vaak af van de keuze die de computerfabrikant voor de VGA-kaart heeft gemaakt. Ook kan de slotruimte in de PC een beperkende factor zijn.
Een 16 bit VGA-kaart is gemiddeld twee maal zo snel als een 8 bit model. Zoals hierboven al vermeld heeft de kaartconnector invloed op de snelheidsaspecten van de VGA-kaart. Deze connector verbindt de VGA-kaart met de systeembus van de PC. De systeembus heeft op zijn beurt ook weer invloed op de VGA-prestaties. Het verschil tussen 8 en 16 bit kan in de praktijk wel eens nauwelijks merkbaar zijn. Een goed ontwerp van de hardware en het gebruik van een recentere driver verrichten wonderen.
Toch speelt ook het busontwerp een belangrijke rol in een succesvol gebruik van de VGA-technologie.
Tussen bijvoorbeeld busontwerpen volgens:
- “Enhanced Industrial Standaard Architecture” (EISA);
- “Micro Channel Architecture” (MCA);
- “Industrial Standaard Architecture” (ISA);

zijn qua snelheid redelijke verschillen te onderscheiden. Vooral in de snellere video-presentatie blinken de geavanceerde EISA- en MCA-bus structuren uit in snelheid en prestatie.

De busklok
Naast de busstructuur is ook de busklok een graadmeter voor de prestaties van een VGA-kaart. Hoe sneller de bus des te sneller ook de video-communicatie. Meestal draait een bus op een kloksnelheid van 8 à 10 MHz, er zijn echter ontwerpen, bijvoorbeeld de flexbus, die sneller kunnen. Vooral de geavanceerde auto-configurerende EISA- en MCA-systemen kunnen profiteren van deze snelle busontwerpen. Veelal worden in dit soort systemen geen standaard VGA-kaarten meer gebruikt. Hier maken vooral de Super-VGA kaarten deel uit van het systeem.

De VGA-connector
Als videoconnector voor de monitor wordt standaard bij analoog VGA de 15 pins sub-D connector gekozen (zie voor de pin-layout figuur 3/19.9-21). Niette-

min zijn er ook weer fabrikanten die 9-pins en BNC-connectors gebruiken. Hetzelfde geldt voor videokaarten waar doorgaans een analoge 15 pins en een TTL 9-pins connector op zitten. In die gevallen zijn de verloopjes en genderchangers onvermijdelijk.



Figuur 3/19.9-21: De aansluitgegevens van de standaard VGA-connector.

Snelheidsparameters van het videogeheugen

Bij het gebruikte videogeheugen (VRAM) kunnen snelle chip's of langzame chip's geïnstalleerd zijn. Zoals reeds vermeld is Video-RAM sneller dan gewoon DRAM, maar ook de refresh-rate van de gebruikte VRAM-chips (70 ns tegen bijvoorbeeld 100 ns) kunnen de beeldopbouw drastisch beïnvloeden. Bij het Video-RAM komt naast de snelheidsparameters van de RAM-chips zelf, ook de gebruikte bandbreedte om de hoek kijken. Deze bandbreedte geeft het aantal bit aan waarmee de VGA-data in de bitplanes uitgewisseld worden. Standaard voor de standaard-VGA was een configuratie van 8 RAM-chips met een bandbreedte van elk 4 bit. Hierdoor vindt de informatie-uitwisseling in 8 x 4 = 32 bit plaats. Om nu 16 kleuren te kunnen adresseren zijn zo 4 onafhankelijke RAM-banken met een bandbreedte van 8 bit nodig. Wil men de bandbreedte van 8 bit naar 16 bit opvoeren, dan vergt dit 16 RAM-chip's. Het resultaat van bandbreedte-vergroting is het aanmerkelijk sneller worden van de VGA-kaart.
De ontwikkelingen in de chiptechnologie hebben echter het gebruik van meerdere 256 kB chip's achterhaald zodat een wijziging van bovengenoemde accesstechniek noodzakelijk is geworden. Veel fabrikanten lossen dit probleem op door memory-management chip's te gebruiken die met 1, 2 of 4 RAM-chips met 16 bit breedte kunnen werken. Een voorwaarde van deze techniek is wel dat het 512 kB of 1 MB videogeheugen als 4 x 256 kB geheugenbanken moet worden geconfigureerd.
Andere versnellende RAM-truc's, die men onder andere ook terug vindt in de PC en die meer en meer in de videotechnologie toegepast worden, zijn videocaching en paging. Deze technieken vindt men echter niet vaak meer in de al wat oudere standaard VGA-kaart maar voornamelijk geïntegreerd in de Enhanced- en Super-VGA kaarten.

Video-coprocessor

Bij de opvolgers van de oer-VGA treft men voor het eerst een techniek aan, die de specificaties van grafische kaarten tot nu toe zullen bepalen: co-processoren. Wan-

19.9 Principes van de beeldscherm elektronica

neer een video-coprocessor door de fabrikant op de kaart is geïnstalleerd, wordt de centrale processor van de PC ontlast. Zonder een dergelijke coprocessor wordt onder andere het adresseren van het beeldschermgeheugen door de PC-processor uitgevoerd. Gedurende deze tijd kan de processor geen andere taken uitvoeren. Bij gebruik van een video-coprocessor wordt de beeldscherm adressering aan deze laatste overgelaten zodat de PC-processor zijn normale taken kan blijven uitvoeren.

De standaard VGA-chip
De VGA-chip bevat vier belangrijke functies, ondergebracht in de verschillende circuits van de chip. Hieronder volgt een korte opsomming van deze functies met een korte uitleg.

- CRT-Controller
  Allereerst is er de "CRT-Controller". Deze verzorgt het aantal gescande lijnen op de monitor, regelt de hardware-interrupts, geeft de timersignalen af en plaatst eventuele animaties achtereenvolgens in het videogeheugen.
- Graphics Data Controller
  Dan is er de "Graphics Data Controller". Deze controller regelt de bewerkingen op de pixels in de vier parallelle bitplanes in het Video-RAM.
- Attribute Controller
  De "Attribute Controller" zet de gewenste kleuren om in de verschillende pixelcodes, vertaalt het attribuutbyte in de tekstmodus en draagt zorg voor het "loze" beeldschermpunt in het 9 dot brede font.
- Sequencer
  De "Sequencer" tenslotte, heeft een scheidsrechtersfunctie bij het geven van toegang aan de verschillende fonts tot het display-RAM. Van de 8 fonts kunnen telkens 2 tegelijk op het beeldscherm actief zijn.

De digitaal analoog convertor
De DAC (ook bekend onder de typeaanduiding INMOS G171), genereert 256 kleuren uit een palet van 256.000 kleuren op een analoge RGB-monitor. Om deze kleuren te kunnen genereren maakt de DAC gebruik van een kleuren opzoektabel. In HIRES-mode zijn dit echter slechts 16 kleuren uit een palet van 256 kleuren. Naast de genoemde kleurenpracht kan de DAC ook een 3D-shading techniek op het beeldscherm brengen. Dit wordt voornamelijk in CAD-toepassingen gebruikt.

Soft-font technologie
In het VGA-tijdperk wordt de argeloze PC-gebruiker geconfronteerd met een nieuwe kreet: soft-font technologie. Deze door ATI ontwikkelde technologie is te vergelijken met de "Enhanced Printing Technology" die voor laserprinters wordt toegepast. Met deze technologieën is het mogelijk het beeldscherm of de printer een schijnbaar hogere resolutie te geven dan deze apparaten in werkelijkheid hebben. Een normale zwarte letter wordt door Windows op het scherm afgedrukt door pixels wit of zwart aan te sturen. Een tussenoplossing is er niet. Daardoor ontstaan de bekende getrapte rafelige omtrekken van de letters, die erg opvallen bij grote corpsen. De soft-font technologie voegt nu aan de randen van de letters nog grijze pixels toe. In totaal staan hiervoor 16 verschillende grijstinten ter beschikking. Het gevolg is dat de omtrek van de letter voor het oog veel ronder wordt en een tekst er op het scherm uitziet alsof afgedrukt op een 300 dpi laserprinter. Hierdoor wordt een schijnbare schermresolutie verkregen van 3.200 x 2.400 pixels.

19.9 Principes van de beeldscherm elektronica



Figuur 3/19.9-22: Vergelijking van de schermafdruk van een pagina uit "Ventura for Windows" zonder soft-font technologie (boven) en mét (onder).

5	4	3	2	1	0
SROO	SGRN	SBLA	PROO	PGRN	PBLA

SROO: secundair rood 1=aan 0=uit

PROO: primair rood 1=aan 0=uit

SGRN: secundair groen 1=aan 0=uit

PGRN: primair groen 1=aan 0=uit

SBLA: secundair blauw 1=aan 0=uit

PBLA: primair blauw 1=aan 0=uit

Figuur 3/19.9-23: De eigenschappen van VGA.

In figuur 3/19.9-22 is een vergelijking gegeven van een scherm zonder en mét deze technologie. Kommentaar overbodig!
De soft-font technologie werkt met eigen fonts. Toch worden ook de fonts van Windows- en de "True Type"-fonts ondersteund. Dank zij de snelheid van de moderne VGA-kaarten worden deze fonts "on the fly", dus op het moment dat zij naar het beeldscherm worden gestuurd, omgezet in soft-fonts.

Voor- en nadelen van de standaard VGA-kaart

De grote voordelen van de standaard VGA-kaart zijn vooral de compatibiliteit met de EGA-standaard en het feit dat vrijwel elk software-pakket een software driver heeft voor deze kaart. Ook zijn de ontwikkelingen in de chip-technologie nu zover gevorderd dat het mogelijk is om de gehele VGA-logica in één enkele chip onder te brengen. Deze techniek komt in de steeds verdergaande miniaturisering van PC's goed van pas. Het gebruik van deze techniek in de laptop en notebook PC's is al eerder genoemd, maar ook in de industriële automatisering kan de VGA-techniek nu worden ingezet.
Een ander voordeel van de standaard VGA-kaart is dat de goedkopere monitoren die geschikt waren voor EGA, ook voor deze standaard voldoen.
De nadelen van de standaard VGA-kaart ten opzichte van de overige twee VGA-categorieën zijn onder andere het beperkte aantal kleuren in de HIRES-mode en het toch nog zwakke oplossend vermogen van 640 bij 480 pixels.

Samenvatting

In figuur 3/19.9-23 zijn de voornaamste eigenschappen van de VGA-standaard samengevat.

19.9 Principes van de beeldscherm elektronica

Mode	Resolutie	tekst/graphics	kleur/mono	Emulatie
0	40x25	tekst	mono	CGA
1	40x25	tekst	kleur	CGA
2	80x25	tekst	mono	CGA
3	80x25	tekst	kleur	CGA
4	320x200	graphics	4 kleuren	CGA
5	320x200	graphics	mono	CGA
6	640x200	graphics	mono	CGA
7	80x25	tekst	mono	MDA
D	320x200	graphics	16 kleuren	EGA
E	640x200	graphics	16 kleuren	EGA
F	640x350	graphics	mono	EGA
10	640x350	graphics	16 kleuren	EGA
11	640x480	graphics	2 kleuren	VGA
12	640x480	graphics	16 kleuren	VGA
13	320x200	graphics	256 kleuren	VGA
14	800x600	graphics	16 kleuren	E-VGA

Figuur 3/19.9-24: Alle VGA-modes inclusief Enhanced-VGA.

EVGA

Inleiding

Naast de oude EGA-modes en de extra modes van de standaard VGA-kaart kent de Enhanced-VGA (ook wel Extra-VGA genoemd) nog een mode in high resolution graphics met een resolutie van 800 x 600 beeldschermpunten (zie figuur 3/19.9-24). Ook hier kunnen dan 16 kleuren uit een palet van 256 kleuren tegelijkertijd op het scherm worden getoond.

Videogeheugen op de kaart

Bijna alle Enhanced-VGA kaarten zijn uitgerust met 512 kB of 1 MB aan Video-RAM geïnstalleerd op de kaart.

Dit is niet verwonderlijk omdat een beeld opgebouwd door deze kaart, toch uit 800 x 600 = 480.000 pixels bestaat. Bij een pixelgrootte van 8 bit is dit een aanslag op het videogeheugen van ongeveer 470 kB.

Voor- en nadelen van Enhanced-VGA

Enhanced-VGA is eigenlijk niet meer dan een uitbreiding op standaard VGA waarbij de maximale resolutie verhoogd is van 640 x 480 pixels naar 800 x 600. Het voordeel van deze 800 x 600 pixelresolutie is dat nu ook grafische figuren, zoals bijvoorbeeld een cirkel, vloeiender worden weergegeven. Toch is een van de belangrijkste nadelen van standaard VGA-kaart ook bij de Enhanced-VGA nog niet verholpen. Ook nu zijn slechts 16 kleuren in HIRES-mode tegelijkertijd op het scherm zichtbaar te maken.

SVGA

Inleiding

De Super-VGA kaarten starten vanaf de resolutie van 1.024 x 786 pixels. Het einde van deze resolutie-vergroting is hiermee echter nog lang niet in zicht! Er komen kaarten op de markt met resoluties van 1.152 x 864 pixels, 1.280 x 1.024 pixels en 1.600 x 1.200 pixels.
Natuurlijk zijn alle Super-VGA kaarten downwards compatibel met VGA, EGA, MCGA en vaak ook nog met CGA. Bij veel VGA-kaarten wordt naast de software-driver ook nog een aantal utilities geleverd om bijvoorbeeld verschillende grafische standaarden te kunnen emuleren.

Interlace

Super-VGA kaarten kunnen worden ingedeeld in "interlaced" en "non-interlaced". De interlace-techniek is een goedkope techniek om met dezelfde hardware meer beeldscherpte te creëren. Dit wordt bereikt door het afzonderlijk optekenen van de even en oneven beeldlijnen. Men spreekt dan van een interlaced display. In de praktijk mag het beeld dan wel scherper ogen maar het resulterende flikkerende beeldscherm is meestal niet om lang naar te kijken.

De resolutie

In de tabel van figuur 3/19.9-25 zijn de "meer dan gewoon VGA"-resoluties van de EVGA- en SVGA-kaarten nog eens samengevat.

Geheugen op de kaart

Het aantal beschikbare kleuren bij VGA was in principe 262.144. In de praktijk speelt het beschikbare video-RAM echter een zeer belangrijke rol. Bij een videogeheugen van 256 kB zijn er in de 800 x 600 mode slechts 16 kleuren ter beschikking. Uitbreiden tot 512 kB of terugschakelen naar 640 x 480 brengt maximaal 256 kleuren tegelijkertijd op het scherm. Bij een resolutie van 1.024 x 768 is er voor hetzelfde display minimaal 768 kB aan videogeheugen nodig. De Super-VGA kaarten werden dus bijna zonder uitzondering met 1 MB aan video-RAM geleverd.

Tekstweergave

Hoe meer pixels des te meer teksttekens er op het monitorscherm passen. Het standaard IBM tekstdisplay bestaat uit 80 tekens en 25 regels.
Bij 1.024 x 786 Super-VGA variëren de maximaal haalbare tekstmodes van 90 tekens bij 30 regels tot 100 tekens bij 40 regels of zelfs 132 tekens bij 44 regels. Het gebruik van een voldoende grote monitor (16 tot 20 inch beelddiagonaal) is

19.9 Principes van de beeldscherm elektronica

dan aan te raden omdat anders de letters te klein uitpakken.
Bij 1.024 x 786 beeldschermpunten lijken de letters scherper dan bij de lagere VGA-resoluties. Dit heeft twee mogelijke oorzaken:
- de letters worden kleiner weergegeven zodat ze meer dichtvloeien en daarom scherper ogen;
- bij geschikte screenfonts bestaan de Super-VGA letters uit meer puntjes en juist daarom ogen zij scherper en vloeiender.

De verbetering van de tekstweergave kan in de praktijk nogal eens tegenvallen. Dit is vaak te wijten aan een verkeerd schermfont, een te klein lettertype, een verkeerde monitor of een ondoelmatige screendriver. Alleen echte Super-VGA-fonts, een compatibele goede monitor en de optimale driver geven de best haalbare beeldkwaliteit.

Bandbreedte van monitoren

Een elektronenkanon (de diverse kleurenmonitoren gebruiken 3 elektronenkanonnen voor de kleuren rood, groen en blauw) wordt voortdurend aan- en uitgeschakeld. De bandbreedte van een monitor geeft aan hoe vaak een elektronenkanon tijdens het scannen aan en uit geschakeld kan worden. Daarmee definieert de bandbreedte hoeveel pixels de monitor per tijdseenheid op een beeldlijn kan zetten. Hoe hoger de bandbreedte in MHz des te hoger is het oplossend vermogen van de monitor. Elke videostandaard kent een minimum bandbreedte.
De meeste monitoren geschikt voor een bepaalde videostandaard hebben een grotere bandbreedte dan voor de standaard is vereist. Normaal zijn bandbreedtes van 35 tot 70 MHz. Voor de VGA-standaard is gezien het oplossend vermogen van 640 x 480 pixels minimaal 34 MHz nodig. Hoewel standaard-VGA nog wel met een EGA-compatibele RGB-monitor werkt is het noodzakelijk om, vanaf de 800 x 600 Enhanced-VGA resolutie, een echte multisync-monitor in te zetten.
Dit type monitor beschikt over de juiste bandbreedte om de horizontale en verticale scanfrequenties te kunnen weergeven. Het grote voordeel van multisync-monitoren is dat de monitor zelf de frequenties herkend die de videokaart afgeeft en op die manier de elektronenbundel (of bundels voor kleurenmonitoren) het scherm laat scannen met de juiste frequentie. Bovendien past deze monitor ook bij de lagere videostandaarden. Bij het grote oplossend vermogen van de 1.024 x 768 Super-VGA kaart, maar ook bij nog hogere resoluties met veel kleuren zijn echter specialistische multisync-monitoren nodig met bandbreedtes tot wel 100 MHz.
Een nadeel van deze grote bandbreedtes is dat wanneer men gebruik wil maken van "live"-TV of videobeelden er een lelijk geflikker op het beeldscherm verschijnt.
Bij VGA passen er namelijk in één beeldseconde 60 tot 70 scancycli. De scanfrequentie van een VGA-monitor bedraagt dus met andere woorden 60 tot 70 Hz (voor E- en S-VGA is dit nog meer!). Daardoor ontstaan problemen voor hen die de PC voor het maken van 50 Hz PAL-videobeelden willen gaan gebruiken. Dan is er speciale dure videohardware nodig om het irritant flikkerende effect tegen te gaan.

De dot-pitch

Bij de hoge resoluties waartoe SVGA in staat is, wordt een tot nu toe volledig onbelangrijke eigenschap van de monitor opeens heel belangrijk: de dot-pitch.

19.9 Principes van de beeldscherm elektronica

Kaart	Resolutie	Mode	Kleuren	HSF	VSF
Enhanced	800x600	graphics	16/256	56 Khz	35,16 Hz
Super	1024x768 interlaced	graphics	16/256	43,4 Khz	35,46 Hz
	1024x768 non-interlaced	graphics	16/256	47,2/57 Khz	58/70 Hz

Figuur 3/19.9-25: Resolutie-eigenschappen E- en S-VGA.

De dot-pitch geeft aan hoe groot (in feite hoe klein) de fosfordeeltjes op het beeldscherm zijn.
Alleen een dotpitch van 0,28 mm of minder en een beelddiagonaal van minimaal 14 inch (liefst 16 of 20) geven een voldoende beeldkwaliteit.

XGA

Inleiding
XGA, afkorting van "Extended Graphics Array", is een grafische standaard waarbij een snelle VGA-kaart en een grafische processor, gecombineerd met bus-mastering, op een insteekkaart geïntegreerd zijn. Bijna vijf jaar nadat de VGA-standaard werd geïntroduceerd kwam alweer IBM met een nieuwe generatie PS/2 grafische hardware, de XGA. In het desktopmodel 90 werd de XGA-logica als eerste op het moederbord geïntegreerd. In het towermodel 95 werd de XGA logica op een aparte Micro Channelinsteekkaart ondergebracht.

XGA als opvolger van de 8514/A adapter
IBM's strategie om de VGA-logica te vervangen door de XGA grafische standaard was opmerkelijk. Enkele jaren voordat de XGA-adapter werd geïntroduceerd waren de geruchten niet van de lucht dat IBM de 8514/A Advanced Graphics technologie op PS/2 moederborden zou gaan integreren. De 8514/A-kaart mistte echter een zeer belangrijke eigenschap die het gebruik van deze kaart zou kunnen rechtvaardigen en dat was de "backward" compatibiliteit. De 8514/A-kaart was niet in staat om voorgaande grafische standaards zoals EGA en VGA volledig te ondersteunen. De XGA-standaard heeft een volledige VGA hardware compatibiliteit en kan dus goed als grafisch platform op een moederbord worden geïntegreerd.

Overzicht
XGA was op een aantal punten een samenvoeging van VGA en 8514/A logica. In de tabel van figuur 3/19.9-26 worden de verschillende kenmerken van XGA vergeleken met VGA en SVGA.

De ontwikkeling van de XGA standaard
De XGA grafische kaart werd net als de 8514/A-adapter en de specialistische Image Adapter/A ontwikkeld in Groot Brittanië in de IBM laboratoria van Hursley. Met deze achtergrond is het niet verwonderlijk dat veel features uit de 8514/A-adapter ook in de XGA terug te vinden zijn, al zijn deze eigenschappen vaak op een andere wijze in de XGA-logica geïmplementeerd.

19.9 Principes van de beeldscherm elektronica

Platform	Resolutie	Interl.	Comp.	Graf. Coproc.	Memory-mapped	RAM type
PS/2 VGA	640x480x16 320x200x256	n.v.t.	VGA	Nee	Ja	DRAM
SuperVGA	800x600x16/256 1024x768x16/256	Nee	VGA	Nee	Ja	DRAM
8514/A	640x480x16/256 1024x768x16/256	Ja	Geen	Ja	Nee	VRAM
XGA	640x480x2/4/16 640x480x256/65536 1024x768x2/4/16/256	Ja	VGA	Ja	Ja	VRAM

Figuur 3/19.9-26: Vergelijking van de eigenschappen van de vier laatste echte grafische standaarden.

Enkele kenmerken zoals bijvoorbeeld het al eerder genoemde Bus-Mastering zijn speciaal ontwikkeld om volledig te kunnen profiteren van IBM's Micro Channel Architecture dat standaard aanwezig was in de meeste PS/2 computers. Andere kenmerken zoals een memory-mapped frame buffer en een hardware cursor bieden een grotere flexibiliteit dan andere bestaande grafische ontwerpen. Dit is een groot voordeel voor software ontwikkelaars die hierdoor niet genoodzaakt worden om ingewikkelde schermroutines te bouwen.

Nog een voordeel voor de software ontwikkelaars is dat IBM dit keer zeer volledige register-specificaties beschikbaar heeft voor de XGA-kaart, iets dat zeker niet het geval was voor de 8514/A-adapter. Hoewel de XGA niet met de volledige set van 8514/A-kenmerken compatible is, wordt ondersteuning voor deze kaart wel via een Adapter Interface gewaarborgd.

Drie verschillende modes

De XGA heeft drie onderscheidbare modes, te weten:

- VGA compatibiliteit;
- 132 koloms VGA compatibiliteit;
- Extended Graphics mode.

De Extended Graphics mode is de meest interessante van deze drie omdat deze mode de mogelijkheden van zeer hoge resolutie en een grote grafische acceleratie biedt.

VGA compatibiliteit

Om een volledige compatibiliteit te kunnen waarborgen voor de VGA standaard, die van origine ook van IBM afkomstig is, heeft IBM veel geleerd van de enorme hoeveelheid VGA-klonen waardoor de VGA-versie in de XGA-kaart een breder datapad heeft gekregen. De VGA-mode bezit daarom, ondanks een 8 bit intern datapad, een externe busbreedte van 32 bit. Naast deze uitbreiding van de bus-

19.9 Principes van de beeldscherm elektronica

breedte bezit de XGA-kaart in VGA-mode ook een intern schrijfcache waardoor de XGA-chip de busdata kan opdelen in dusdanige databrokjes dat de rest van het systeem niet door onnodige wachtcycli (waitstates) wordt opgehouden.
Volgens IBM's technische documentatie is de XGA-kaart in VGA-mode 90 % sneller als de originele VGA-kaart onder het MS-DOS operating systeem en ongeveer 50 % sneller in de DOS-Windows combinatie.
Behalve de performance verbeteringen zijn er geen veranderingen in de VGA functionaliteit.

Meer XGA adapters in één systeem
Hoewel het fysiek mogelijk is een computersysteem met maximaal acht XGA-kaarten te configureren (de CMOS configuratie software kan slechts zes XGA-kaarten aansturen) kan er slechts één XGA-kaart in VGA-mode actief zijn. Is dit niet het geval dan kunnen er I/O-konflikten in het systeem optreden waardoor de computer crasht.

132 koloms VGA
Bij het gebruik van de 132 koloms mode (een VGA-extensie) is het mogelijk om 132 karakters per regel te tonen en te manipuleren.
De karakterbreedte is 8 pixels bij een horizontale resolutie van 1.056. De karakterhoogte hangt af van het gekozen font. Dit betekent dat men kan kiezen uit tekstresoluties van 132 bij 43, 132 bij 50 of 132 bij 60 pixels. Helaas kon men bij de meeste XGA-kaarten deze 132-koloms mode alleen instellen door handmatig de XGA-registers te manipuleren. Naast het handmatig instellen wordt de mode geselecteerd bij het omschakelen naar de grafische mode, hexadecimaal 14.

Opmerking
Omdat beide VGA-modes alle originele VGA-kenmerken bezitten zijn ook alle beperkingen van deze modes nog in de VGA-modes van de XGA aanwezig.

Extended graphics mode
De extended graphics mode had veel voor die tijd spannende opties zoals bijvoorbeeld 65.536 kleuren bij een resolutie van 1.024 bij 768 beeldpunten, bus mastering, beeldopbouw acceleratie en de al eerder genoemde hardware cursor. Hoewel een aantal van deze opties ook beschikbaar is in de overige modes is het gros van de XGA-registers en -functies bestemd voor de extended graphics mode.

Het XGA-ontwerp
Het XGA ontwerp bestaat uit video-RAM, tweeledig adresseerbaar RAM voor gebruik in grafische display systemen en (de twee gebruikelijke chip's op grafische platforms) een grafische coprocessor en de display controller.
Deze laatste twee vormen het hart van de XGA-kaart. De grafische coprocessor regelt de VGA-compatibiliteit, de tekenfuncties en het videogeheugen beheer. De display controller bevat een RAM digitaal/analoog converter (RAMDAC) met een kleuren "lookup" tabel, de CRT-controller, de hardware cursor ondersteuning en een zogenaamde VRAM-serializer. Deze VRAM-serializer bestaat uit video-logica dat zorg draagt voor het omzetten van VRAM-data naar beeldscherminformatie.

De XGA-registers
Toegang tot de XGA-kaart wordt geregeld door twee sets registers. De eerste set is terug te vinden in het systeem I/O-geheugengebied terwijl de andere set in

19.9 Principes van de beeldscherm elektronica

het RAM-geheugen te adresseren is. De adressen van beide registersets kunnen afhankelijk van de XGA-configuratie variëren. Het bestaan van variabele register adressen geeft meteen aan waarom meerdere XGA-kaarten naast elkaar in één computersysteem kunnen worden geplaatst.
De I/O-registers zijn terug te vinden op de adressen 21X0 hexadecimaal. Hierin geeft "X" een soort basiswaarde of "instance" aan. Volgens IBM heeft "X" de waarde 6 in systemen met slechts één XGA-kaart aan boord, dit resulteert dus in een I/O-register adres van 2160 hexadecimaal.

RAM-registers
De registers die zich in het RAM-geheugen bevinden beslaan een geheugengebied van 128 bytes en wel in het laatste kilobyte van een 8 kB groot geheugensegment. De plaats van een dergelijk 8 kB geheugensegment ligt ergens in de range van de adressen 0C0000 en 0DFFFF hexadecimaal. Het nut hiervan is dat de eerste 7 kB kunnen worden opgevuld door XGA ROM-data. Dit is echter alleen nuttig bij een XGA Display Adapter/A (een adapter op een insteekkaart).
De XGA moederbord uitvoering maakt gebruik van ROM dat op het moederbord-ROM is gemonteerd, voor bijvoorbeeld de XGA initialisatiecode.
Het XGA instancenummer bepaalt de plaats van de 128 bytes binnen het 8 kB geheugensegment.
De I/O-registers beïnvloeden hoofdzakelijk de display controller terwijl de geheugenregisters meer de grafische coprocessor beïnvloeden.
Bij het opstarten van de computer bepaalt de zelftest routine van de XGA de instancenummers voor beide registersets. Door nu het PS/2 POST-register te raadplegen kan men de adressen van de registersets achterhalen voor de betreffende XGA-kaart. Dit laatste geldt dan voornamelijk bij multi-XGA systemen.

Registergrootte
Omdat de XGA-kaart ontworpen is voor 486-systemen zijn de adressen van de meeste geheugen gebaseerde registers 32 bit groot. Naast de Intel adresseringsmethodiek heeft IBM de XGA-kaart ook adresseerbaar gemaakt voor gebruik in Motorola's 68000-omgevingen.

Initialisatie
Initialisatie is een noodzakelijke stap in het gebruik van grafische adapters. In het geval van de XGA-adapter bestaat de initialisatie voornamelijk uit het instellen van de extended graphics mode via het operating-mode register. Daarna is het mogelijk om via het CRT-controleregister de gewenste resolutie te selecteren. De selecteerbare resoluties zijn 640 x 480 pixels in 65.536 kleuren en 1.024 x 768 pixels in 256 kleuren. De resoluties in kleur zijn echter alleen beschikbaar wanneer de XGA-adapter met 1 MB aan video-RAM geheugen (VRAM) is uitgerust.

16 bit per pixel in 65.536 kleuren
De 16 bit per pixel mode (met 65.536 kleuren) geeft een beeld dat het kleurenspectrum van een foto aardig benadert, dit betekent dat een foto die met behulp van een scanner is ingelezen, op een XGA beeldscherm bijna een exacte replica geeft van het originele plaatje. De samenstelling van een 16 bit pixel in kleur bestaat uit 5 bit rood, 6 bit groen en 5 bit blauw (5-6-5). Dit zijn met andere woorden 32 blauwe, 32 rode en 64 groene tinten in ieder pixel. Vormen van dergelijk configuraties zijn terug te vinden in de

19.9 Principes van de beeldscherm elektronica

PC-standaard TARGA met een 5-5-5 samenstelling en het i860 formaat met een 6-6-4 configuratie. Blijkbaar is het menselijk oog meer gevoeliger voor groene tinten dan voor rood of blauw. De reden dat voor de kleuren rood, groen en blauw gekozen is, is dat alle gangbare monitoren een rood, groen en blauw elektronenkanon bezitten. Alle overige kleuren kunnen met deze basiskleuren worden samengesteld.

De display-controller

Naast het gebruik van de display-controller tijdens de initialisatie, kan de controller bijvoorbeeld ook worden ingezet bij acties op de kleuren opzoektabel (lookup-tabel) of bij het gebruik van sprites.

Lookup-tabel

De lookup-tabel wordt gebruikt om de 1, 2, 4 of 8 bit pixelwaardes te vertalen in de bijbehorende RGB-waardes. De pixelwaarde fungeert hierbij als een index in de lookup-tabel. De RGB-waardes worden vervolgens geconverteerd van de digitale niveau's naar analoge spanningen door middel van een ingebouwde digitaal/analoog converter (DAC).
Net als bij de VGA- en 8514/A-adapters ondersteunt de XGA kleuren opzoektabel 64 niveau's (6 bit) voor iedere primaire kleur met een totaal van 262.144 mogelijke kleuren combinaties. Dit betekent bijvoorbeeld voor de 256-kleuren mode dat men kan kiezen uit 256 kleuren en die weer uit een kleurenpalet van 262.144 tinten.

Sprites

Een sprite is een 64 bij 64 pixelblok dat over het bestaande scherm kan worden geplaatst. Ieder pixel in een sprite heeft vier mogelijke waardes; sprite kleur 0, sprite kleur 1, transparantie en complementeren.
Speciale registers definiëren de sprite kleuren en specificeren de RGB-waardes voor iedere kleur. De RGB-waardes worden direkt naar de DAC gestuurd. Het gebruik van sprites geeft applicaties bijvoorbeeld de mogelijkheid een cursor, onafhankelijk van gekleurde delen op het scherm, over het scherm te laten bewegen. Transparantie biedt de mogelijkheid dat niet alle pixels in het 64 x 64 blokje ingevuld moeten worden. Met behulp van het complementeren kan een sprite altijd over ieder beeldschermdeel zichtbaar zijn.

Task switching

Een van de grootste problemen die systeemsoftware ontwikkelaars bij het schrijven van multi-tasking omgevingen hebben, is het veilig stellen van de toestand van de grafische adapter, zodat een andere applicatie de grafische kaart kan overnemen. Deze methode van veilig stellen van de toestand van de grafische kaart moet ook rekening houden met het feit dat de kaart bezig kan zijn met beeldschermopbouw, of dat het kleurenpalet wordt bijgewerkt. De XGA-kaart werd duidelijk ontwikkeld met task switching in gedachte, er zijn een flink aantal faciliteiten toegevoegd voor het veilig stellen en terugbrengen van de toestand van de grafische kaart.

Het definiëren van de tekenruimte

Een van de unieke mogelijkheden van de XGA-kaart is het gebruik van bitmap's die kunnen worden gedefinieerd voor het uitvoeren van tekenfuncties.
Deze bitmap's bestaan uit lineaire delen van het geheugen die gedefinieerd zijn

19.9 Principes van de beeldscherm elektronica

met een pixelhoogte, pixelbreedte en een pixeldiepte. Dit betekent dat een 8 bit per pixel bitmap met een breedte van 10 en een hoogte van 6 pixels 60 bytes geheugen beslaat. Het laatste pixel of byte uit een beeldschermregel grenst in het geheugen direkt aan het eerste pixel of byte van de volgende beeldschermregel. Anders dan bij andere grafische kaarten kan het bitmapgeheugen overal in het adresseerbare systeemgeheugen worden gesitueerd. Dit betekent dat grafische data niet meer van het werkgeheugen naar het XGA VRAM-geheugen behoeft te worden gekopieerd om een beeld op het scherm te kunnen tonen. Normaal bevindt het VRAM-geheugen zich in het hoogste adresseerbare gebied van de 4 GB adresruimte die theoretisch aanwezig kan zijn in moderne computersystemen. Dit betekent concreet dat er geen geheugenconflict kan optreden.

Valkuilen van bus-mastering
Wanneer de XGA-kaart een bitmap adresseert, dan wordt eerst gecontroleerd of het hier om een adres in het locale geheugen (VRAM) gaat of dat het adres zich in het "remote" geheugen (systeemgeheugen) bevindt. Voor toegang tot het "remote" geheugen moet de XGA-kaart gebruik maken van de systeembus om het systeemgeheugen te kunnen adresseren. Hier komt dan de XGA bus-mastering methode om de hoek kijken. Naast de processor-opdrachten die nodig zijn om het locale geheugen te kunnen adresseren is er een extra overhead aan processor-opdrachten nodig om het systeemgeheugen te kunnen adresseren.
Daarnaast bestaat er het performance voordeel wanneer de XGA-processor het geheugen manipuleert zodat de systeemprocessor andere taken kan blijven uitvoeren. Toch heeft deze bus-mastering methode enkele nadelen.
Zowel de 80386- als de 80486-processor ondersteunen een techniek van virtueel geheugen beheer door middel van een interne pagina map-tabel. Deze pagina map-tabel geeft de zogenaamde controle applicaties zoals expanded-memory managers, DOS-extenders en geavanceerde operating systemen de mogelijkheid om virtuele PC-adressen van 4 kB per stuk (ofwel één geheugenpagina) aan te maken. Dit betekent bijvoorbeeld voor een stukje software dat data naar een adres XXXXX schrijft, dit in werkelijkheid (vertaald door de pagina map-tabel) naar adres YYYYY schrijft. In vele gevallen kan een applicatie echter onmogelijk weten of een adres nu virtueel of juist fysiek is. Om bus-mastering "netjes" te kunnen gebruiken heeft de XGA-kaart een fysiek adres nodig, anders zou data van en naar verkeerde adressen gekopieerd kunnen worden, met desastreuze gevolgen. In enkele controle software omgevingen zijn verschillende mogelijke oplossingen voor dit dilemma geïmplementeerd.
Wanneer een applicatie toegang heeft tot zo'n controle software pagina map-tabel, dan kan het programma informatie doorsturen naar de XGA-kaart. Deze maakt op zijn beurt weer gebruik van de pagina map-tabel zodat er, virtueel naar fysiek, adres-translaties kunnen worden uitgevoerd.
Helaas leveren veel controle programma's en operating systemen geen toegang tot deze pagina map-techniek. In ieder geval hebben normale DOS-georiënteerde applicaties nog de beste kans om te kunnen profiteren van bus-mastering omdat het standaard 640 kB werkgeheugen zelden of nooit virtueel zal worden gemaakt.

19.9 Principes van de beeldscherm elektronica

Banking
Naast het altijd herschikken van het VRAM in het hoge geheugen kan, met behulp van software, de XGA-kaart in een 64 kB groot geheugensegment op adres A0000 Hex of B0000 Hex worden ondergebracht. Dit geheugensegment is het standaard PC video-geheugen. Wil men nu andere 64 kB geheugensegmenten in het XGA-VRAM op dezelfde manier selecteren, dan hoeft slechts een andere indexwaarde in het (aperture) indexregister te worden geschreven. Dit zogenaamde "banking"-mechanisme is heel handig in real-mode applicaties. Natuurlijk kan ook het volle 1 MB VRAM in een geheugensegment ergens in de 16 MB grote adresruimte worden geplaatst mits er geen geheugenconflict optreedt.
Dit type van geheugenmapping is bij voorkeur geschikt voor protected-mode applicaties.

Bitmap's
De XGA-kaart heeft drie vrij definieerbare en voor algemeen gebruik geschikte bitmap's: map A, B en C. Aan deze bitmap's wordt gerefereerd wanneer tekencommando's worden uitgevoerd. De tekencommando's hebben een of meer van deze bitmap's nodig om onder andere de bron (source bitmap), de bestemming (destination bitmap) en het patroon (pattern bitmap) van de tekenactie te kunnen berekenen. In de bron of source bitmap's bevindt zich de data die, ofwel gekopieerd ofwel rechtstreeks, gebruikt kan worden. De bestemming bitmap, ook wel destination bitmap genoemd, wordt gebruikt om data naar te kopiëren, of om data in te tekenen. Patroon of pattern bitmap's tenslotte, worden gebruikt om pixelverzamelingen in een bepaald patroon te kunnen weergeven.

Maskeren
Naast de bron, bestemming en patroon bitmap's, ondersteunt de XGA-kaart nog een vierde bitmap, de zogenaamde masker bitmap. Deze kan gebruikt worden om delen van het beeld af te kunnen kappen. Deze techniek wordt ook wel "clipping" genoemd. Dit gaat als volgt in zijn werk. Ieder bit met waarde 0 in het masker kenmerkt een pixel dat bij een tekenactie niet mag worden gemanipuleerd, terwijl een bit met waarde 1 aangeeft dat het betreffende pixel wel mag worden bewerkt. Deze techniek wordt handig gebruikt in venster-georiënteerde omgevingen. Overlappende vensters kunnen met behulp van de masker bitmap onafhankelijk van elkaar op het beeldscherm worden gemanipuleerd. Bij een vol scherm beslaat de corresponderende masker bitmap in een 1.024 x 768 resolutie slechts 97 kB videogeheugen.

Pixel map registers
Alle hierboven genoemde bitmap's worden via 5 zogenaamde pixel map registers gedefinieerd. Het eerste pixel in iedere bitmap bevindt zich in de linker bovenhoek en heeft de coördinaten 0,0. Het masker map origin register definieert de positie van het masker over het originele beeld. Alle map's zijn in hoogte en breedte gelimiteerd tot 4.096 bij 4.096 pixels.

Tekenen met de XGA
Voordat men ook maar iets kan tekenen met behulp van de XGA-kaart moet tenminste de bestemming map (destination map) van de gekozen operatie gedefinieerd zijn. Bij bewerkingen als bijvoorbeeld de BitBlt-operatie moet naast een bestemming map ook een bron map (source map) gedefinieerd zijn. Een BitBlt-bewerking kopieert pixels van één

plaats van het scherm naar een andere plaats.
De beschikbare tekenbewerkingen van de XGA zijn:
- lijnen;
- korte vectoren;
- patroongevulde rechthoeken;
- BitBlt's;
- patroongevulde schermgebieden.

Helaas is het weergeven van een simpele lijn met een XGA-kaart niet opgelost door slechts twee coördinatenparen in te geven. De XGA-kaart gebruikt een Bresenham lijnen algoritme. Om dit algoritme echter te kunnen gebruiken moeten eerst de initiële Bresenham parameters worden berekend. Dit betekent een flinke overhead aan berekeningen voor het tekenen van een eenvoudige lijn. Deze methode van lijnen tekenen is rechtstreeks afkomstig van de XGA voorganger, de 8514/A-adapter.
De XGA korte vectoren lijken net als het bovengenoemde lijnen algoritme sterk op de uitvoering van korte vectoren in de 8514/A-adapter. Deze vectoren kunnen maximaal 15 pixels lang zijn en in veelvouden van 45 graden zowel horizontaal, vertikaal en diagonaal gericht worden. Het voordeel van de XGA korte vectoren is dat deze slechts 1 byte aan geheugen vullen. In de 32 bit omgevingen van de moderne PC's kunnen 4 bytes tegelijkertijd worden verwerkt zodat de korte vectoren in een snel tempo kunnen worden gemanipuleerd en dat resulteert weer in een hoge teken snelheid.
Het tekenen van rechthoeken is zeer eenvoudig in de XGA-logica geïmplementeerd, het specificeren van positie, hoogte en breedte is al voldoende om een rechthoek op het scherm te laten verschijnen.
De al eerder genoemde BitBlt-bewerking is afgeleid van de rechthoek bewerking, behalve dat men bij de BitBlt-bewerking ook nog het brongebied moet aangeven waaruit de pixelgegevens moeten worden gekopieerd. Daarnaast is het met de BitBlt-techniek ook mogelijk om vlakken van kleur te doen veranderen door bron- en bestemmingsgebied gelijk te kiezen maar bij het kopiëren (naar dezelfde positie) een andere kleur aan de pixels te geven.
Het vullen van gebieden met een bepaald patroon is natuurlijk ook een afgeleide van de techniek om rechthoeken te tekenen en eventueel met een patroon te vullen. Bij het vullen van een gebied gebruikt de XGA grafische coprocessor de patroon map als richtlijn voor de scanconversies naar het beeldscherm. Deze vultechniek gebruikt daarvoor bitwaardes in de patroonmap om aan te geven of een beeldschermpixel het vooraf gedefinieerde patroon krijgt of niet. Initieel zijn de patroonbits uitgeschakeld voor iedere scanregel.
Naast gebieden kunnen ook lijnen, korte vectoren en rechthoeken gebruik maken van de bron- en patroonmap. De patroonmap kan bijvoorbeeld ingezet worden voor lijntypes en gebiedspatronen, de bronmap op zijn beurt voor het roteren en verplaatsen van beeldschermgedeeltes of om patronen van een (andere) kleur te voorzien.

Verdere factoren

Alle tekenfuncties kunnen verder nog beïnvloed worden door een viertal factoren, namelijk kleur, raster operaties (ook wel mixes genoemd), kleurvergelijkingen en pixel bit maskeren.
Het tekenen in kleur is eenvoudig, het is slechts een kwestie van het definiëren van een voorgrond en achtergrond kleur. Bij raster operaties of mixes worden de pixels

19.9 Principes van de beeldscherm elektronica

in het brongebied en de pixels van de bestemming als het ware "gemixed" met bijvoorbeeld een exclusive-OR (XOR) bewerking. Voorbeelden van deze mixtechniek met de XOR-bewerking zijn de tekstcursor en de oplichtende karakters op het beeldscherm.
De kleurvergelijkingstechniek wordt gebruikt om te bepalen of een pixel tijdens een scancyclus moet worden ververst of niet. Dit wordt bepaald door de kleur van het pixel op het scherm te vergelijken met de kleur van het corresponderende pixel in de bronmap. Wanneer het resultaat van deze vergelijking de waarde "waar" (true) heeft, hoeft het pixel niet te worden ververst.

Voor deze kleurenvergelijking kunnen acht condities worden opgebouwd. Altijd waar (true), altijd onwaar (false), groter dan (>), kleiner dan (<), gelijk aan (=), ongelijk aan (<>), kleiner gelijk aan (≤) en groter gelijk aan (≥).
Het vergelijken van kleuren kan gebruikt worden om te voorkomen dat pixels worden ververst in grafische objecten die zich op de voorgrond van het beeldscherm bevinden.
Het pixel bit masker tenslotte bepaalt welke bits in een pixel gemanipuleerd kunnen worden. Deze mogelijkheid wordt het meest gebruikt bij het beschermen van binaire kleurenreeksen en kleurenvlakken op het beeldscherm.

(wordt vervolgd)

3/19.10
De opstart-procedure van de PC

Inleiding

Traag opstarten

Wie zijn of haar computer (type PC) aanzet, zal opmerken dat het een hele tijd duurt alvorens het systeem bedrijfsklaar is. Dit heeft voor een deel te maken met de belachelijk lange opstarttijd van moderne bedrijfssystemen als Windows 95, maar daarnaast wordt dit verschijnsel verklaard doordat de PC, nadat men de netspanning heeft aangezet, een heleboel acties uitvoert die voor de gebruiker/ster grotendeels verborgen blijven.

Deze acties worden ondernomen door het zogenoemde “BIOS”, afkorting van “Basic Input Output System”. Deze software is permanent in het geheugen aanwezig en wel in het adresgebied $E0000_{HEX}$ tot en met $FFFFF_{HEX}$.

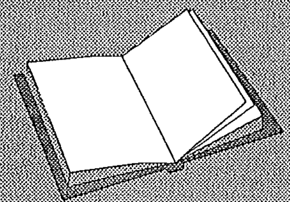
Een belangrijk onderdeel van de BIOS-activiteit is een uitgebreide zelftest, “POST” genoemd. De computer test de belangrijkste onderdelen van zijn elektronisch systeem, onderzoekt hoeveel geheugen in het totale adresseerbereik van de adresbus aanwezig is en welke perifere apparaten zoals harde schijven en diskette-eenheden op het systeem zijn aangesloten. Indien noodzakelijk geeft de opstartroutine de noodzakelijke instructies om ook deze perifere apparaten te initialiseren.

De BIOS-routine bevat ook een subroutine, die eerst schijfeenheid A: in werking stelt en op zoek gaat naar de verborgen start-files van het bedrijfssysteem DOS. Worden deze niet aangetroffen, dan schakelt het systeem automatisch om naar de eerste harde schijf (C:) en gaat hier op zoek naar deze bestanden. Nadat deze zijn ingelezen wordt het interne deel van DOS geladen, de zogenaamde COMMAND.COM. Dit is de gebruikersinterface van DOS die de gebruiker/ster de mogelijkheid biedt gegevens naar de schijven te sturen of files op te slaan.

De verschillende stappen

Het BIOS-systeem voert dus onderstaande stappen uit na het inschakelen van de netspanning:
- genereren van het “Power Good”-signaal;
- uitvoeren van de POST;
- systeemsoftware laden.

LEES OOK:
Hoofdstuk 3/8.9.4
Hoofdstuk 3/19.1
Hoofdstuk 6/10.20



Genereren van het "Power Good"-signaal

Inschakelen van het systeem

Als men een PC aanzet, dan ziet men eerst het scherm even opflitsen, daarna hoort men wat lawaai uit de computer komen (of als men een "stille" computer heeft niet) en als laatste komt er een piepje dat voor de meeste mensen aangeeft dat de computer gebruikt kan worden. Om dit hele proces te starten hoeft men alleen maar één handeling te verrichten, namelijk het aanzetten van de computer. Hierdoor wordt de ingebouwde voeding verbonden met de netspanning. De voeding zal uit de 220 V wisselspanning van het net een aantal gelijkspanningen afleiden, namelijk +5 V, +12 V en -12 V. Deze gelijkspanningen zijn noodzakelijk voor het voeden van de elektronische schakelingen van het systeem. Als eerste zal het systeem een aantal zelftesten uit gaan voeren om te controleren of alle spanningswaarden die de voeding genereert binnen de gestelde grenzen liggen. Deze grenzen zijn vastgelegd in de tabel van figuur 3/19.10-1.

+Pool	-Pool	Minumum	Nominaal	Maximum
P9 - 4,5,6		4.8 V	5 V	5.2 V
P8 - 3		11.5 V	12 V	12.6 V
	P9 - 3	4.5 V	5 V	5.4 V
	P8 - 4	10.8 V	12 V	12.9 V

Figuur 3/19.10-1: De grenzen van de verschillende voedingsspanningen.

De kolommen "+Pool" en "-Pool" geven de pennen van de connector aan waarmee respectievelijk de positieve en negatieve polen van de spanningen verbonden zijn. De term "P9 - 4,5,6" geeft aan dat het hier connector P9 betreft en wel de pinnen 4, 5 en 6.
Er zijn op de voedingsconnectoren P8 en P9 ook verschillende massapunten aanwezig om als referentie te gebruiken. Deze massa wordt aangeduid met de term "Ground". Voor connector P8 zijn dit de pinnen 5 en 6 en voor connector P9 de pinnen 1 en 2. Als massapunt kan men ook de kast van de computer gebruiken.

"Power Good"-signaal

Als de spanningswaarden van de voeding binnen de vastgestelde grenzen liggen, dan wordt er een 5 V gelijkspanningssignaal afgegeven op pin 1 van connector P8 (voeding). Dit signaal staat bekend als het "Power Good"-signaal.
Het kan voorkomen dat het "Power Good"-signaal naderhand weer uitvalt, namelijk als een spanning van de voeding op een gegeven moment buiten de gegeven grenzen valt. Het resultaat hiervan is dat de computer stopt.

Reset naar de processor

Als het "Power Good"-signaal geactiveerd is, dan zal de Timer Chip een reset puls naar de processor sturen, waardoor deze opgestart wordt en begint met het uitvoeren van de "POST".

Uitvoeren van de "POST"

Inleiding

De "POST" is een opstartprocedure van het BIOS die de hardware van het systeem test voordat deze gebruikt gaat worden. POST is een afkorting van "Power On Self Test" en als men dit letterlijk vertaalt, dan krijgt men "Spanning Aan Zelf Controle",

19.10 De opstart-procedure van de PC

hetgeen een heel duidelijke verklaring is van de functie van deze POST!

Functies van POST
De POST voert de volgende testen uit:
- ROM BIOS test;
- Device Init;
- detectie van warme of koude start;
- acoustisch signaal geven;
- systeemsoftware laden;
- klok uitlezen.

ROM BIOS test
In deze test wordt gekeken op welke locatie(s) er een BIOS (Basic Input Output System) beschikbaar is. De standaard adressen voor een BIOS gaan van adres E000:0000 tot en met adres F000:FFFF.

Initialisatie devices
Op dit punt aangekomen wordt de initialisatie van de verschillende devices uitgevoerd:
- initialisatie DMA-controller;
- basis RAM-test;
- initialisatie interrupt controller;
- test moederbord;
- initialisatie graphic controller;
- test interrupt controller;
- initialisatie programmeerbare interval timer;
- test toetsenbord;
- beslissing wel/niet geheugen testen;
- test I/O-gebied;
- test geheugen;
- test hard disk controller.

Procedure
De DMA (Direct Memory Access) controller heeft als functie het transport van gegevens rechtstreeks te laten gebeuren en dus niet via de processor te laten lopen. Op deze wijze wordt het transport aanzienlijk versneld en de belasting van de processor neemt af. De POST test de werking van deze schakeling en voert nadien de Basis RAM test uit. De basis RAM test dient ervoor om de eerste 64 kB van de RAM-banken te testen. De reden hiervoor is dat de POST variabelen gebruikt die in dit geheugen geplaatst worden (RAM-bank 0). Als er dus problemen met deze bank zijn, dan kan de POST zijn gegevens niet kwijt en kan dus niet verder gaan.
Nadien wordt de interrupt controller geïnitialiseerd. Deze chip zorgt ervoor dat de verschillende aanvragen voor computertijd afgehandeld worden. De afhandeling gebeurt op basis van het stellen van prioriteiten, genummerd vanaf nul tot en met vijftien. De allerhoogste prioriteit is nul, hetgeen betekent dat als deze interrupt geactiveerd wordt, de computer altijd deze interrupt eerst af zal handelen. Mogelijke interrupts zijn aanvragen van de timer, het keyboard of een poort.
Bij het testen van het moederbord wordt bepaalde informatie uit het MOS-RAM uitgelezen, zodat verder op in de procedure bepaalde hardware geïnitialiseerd kan worden.
Dit MOS-geheugen wordt door een kleine batterij onder spanning gehouden, zodat de opgeslagen gegevens niet verloren gaan.
Nadien test de POST de aanwezigheid van de grafische processor en voert een paar basis-tests uit. Als laatste onderdeel van deze tests wordt nog de cursor zichtbaar gemaakt op het beeldscherm.
De test van de interrupt controller betekent niets meer dan dat alle interrupts eenmaal softwarematig aangeroepen worden om te kijken of de controller hierop reageert.
Nadien volgt de initialisatie van de programmeerbare interval timer. De functie van deze chip is de ondersteuning van

19.10 De opstart-procedure van de PC

klok-kritieke handelingen. Er zijn drie van deze handelingen, te weten:

- RAM-refresh, het dynamisch werkend RAM-geheugen kan namelijk niet onbeperkt data bewaren en het is dus noodzakelijk op gezette tijden het geheugen eerst uit te lezen en vervolgens weer met dezelfde data te beschrijven.
- Systeemklok, deze functie is zeker klokkritiek.
- Speaker. In eerste instantie kan men zich afvragen of de speaker wel klokkritiek is en zelfs of er wel een klok voor nodig is. Het antwoord hierop is ja, want een computer kan verschillende pieptonen genereren met verschillende frequenties, denk hierbij aan de verschillende audio foutmeldingen.

Nadien zal de POST onderzoeken of er een keyboard aan de computer aangesloten is en of dit een correct toetsenbord is. Er bestaan namelijk verschillende versies! Als dit het geval is, dan dient de keyboardcontroller gereset te worden, omdat in het keyboard-IC altijd data zal staan (ook al heeft deze data geen betekenis) die het systeem in de war kan sturen.
Op dit punt aangekomen dient de POST de beslissing te maken of de volgende geheugen testen wel of niet uitgevoerd moeten worden. Een koude start krijgt men als men de computer aanzet met behulp van de schakelaar en een warme start als men de computer reset als deze al aanstaat (met behulp van de toetscombinatie Alt-Ctrl-Del of met een aparte reset toets).
Bij een warme start moeten deze procedures niet uitgevoerd worden en bij een koude start wel. De POST leest de adressen 0000:0473 en 0000:0472 uit. Als op deze adressen respectievelijk de code 12 en 34 staat, dan is er sprake van een warme start.

Bij de test van het expansion I/O kan men denken aan cartridges die in de computer geplaatst zijn zoals bijvoorbeeld een Bernoulli-drive. Deze cartridges hebben vaak een eigen stukje geheugen. Bij deze test wordt de aanwezigheid van dit type extra geheugen automatisch gedetecteerd en gecontroleerd.
Nadien wordt het extended en expanded memory van de computer getest.
Vervolgens wordt de hard dis controller getest en alle aanwezige diskette stations en hard disk drives getest. De test die hier uitgevoerd wordt is het uitlezen van data op een bepaalde locatie (het lezen of de Boot-files aanwezig zijn). Als deze er zijn worden ze vervolgens uitgevoerd. Het DOS Boot-record dient op cilinder 0, kop 0, sector 1 van de actieve Boot-disk(ette) te staan. De actieve Boot-disk(ette) kan drive A zijn of de eerstvolgende hard disk (dus nooit drive B).

Acoustisch signaal geven
Op dit punt aangekomen zal de POST de overbekende pieptoon laten horen, ten teken dat alle aanwezige hardware getest en goed bevonden is. Dit moet één enkele piep zijn, want als er meerdere pieptonen (van korte of lange duur) komen dan is er iets fout met het systeem.

Systeemsoftware laden
Voor het correct opstarten en functioneren van de computer onder MSDOS dient men de beschikking te hebben over de files IO.SYS en MSDOS.SYS. Deze beide files, die "Hidden" (niet zichtbaar bij een DIR) in de root van de computer staan, zorgen ervoor dat het systeem zonder problemen kan functioneren. Beide files hebben echter verschillende functies. IO.SYS zorgt voornamelijk voor het regelen van de I/O (Input/Output) tussen de ver-

schillende hardware-onderdelen en randapparaten. De functies vallen uiteen in de volgende onderdelen:
- Bepaal de status van de aanwezige hardware.
- Initialiseer de hardware, inclusief de randapparatuur.
- Stel het Disk Operating System in.
- Stel de benodigde systeem parameters in.
- Roep vervolgens MSDOS.SYS aan.

MSDOS.SYS zorgt voor het beheer van de communicatie die van en naar de disk(s) gaat. Tevens worden de verschillende opdrachten die in CONFIG.SYS staan opgehaald en uitgevoerd. De file CONFIG.SYS wordt gebruikt om verdere systeem parameters in te stellen. Hierbij valt te denken aan de volgende mogelijkheden (en er zijn nog meer beschikbaar):
- Instellen van het maximale aantal geopende files en buffers.
- Bepaling van de command processor (in de meeste gevallen COMMAND.COM, maar deze kan ook vervangen worden door een andere command processor).
- Bepaling van de aanwezige device drivers.
- Instelling van de landcode.

Als in de file CONFIG.SYS geen aparte command processor gedefinieerd is, dan is de file COMMAND.COM de zogenaamde primaire command processor. De functie van de command processor is om ervoor te zorgen dat de verschillende (systeem) commando's correct uitgevoerd worden.

Als laatste onderdeel van het opstarten zal COMMAND.COM kijken of de file AUTOEXEC.BAT aanwezig is. Deze file dient ervoor om gebruikers-gebonden instellingen te definiëren en in te stellen. Hierbij valt bijvoorbeeld te denken aan:
- Het instellen van de prompt.
- Instellen van het data-pad met behulp van het Appendcommando.

Klok uitlezen

De klok wordt uitgelezen om de actuele datum en tijd in te stellen. Als het systeem geen geldige data vindt, dan zal de datum standaard op 1-01-80 (1 Januari 1980) en de tijd standaard op 12 uur ingesteld worden.

I/O- en interrupt-adressen

Inleiding

Men kan zich de vraag stellen hoe de POST weet welke onderdelen aanwezig zijn en hoe deze aangesproken kunnen worden voor de tests. Iedere PC maakt hiervoor gebruik van twee belangrijke gegevens, namelijk de I/O-adressen en de interrupt-adressen.

I/O-adressen

De term "I/O" staat voor "input/output". Iedere processor heeft een I/O-adresruimte, waarmee het mogelijk is toegang te krijgen tot bepaalde perifere apparatuur. Deze adresruimte staat volledig los van de adresruimte die wordt gebruikt om het geheugen aan te spreken. Stel dat een 8 bit woord moet ingelezen worden in een bepaald register op een uitbreidingskaart. De processor zet het I/O-adres van dit register op de adres-bus en het woord dat moet ingelezen worden op de databus. Om een duidelijk onderscheid te maken tussen een geheugenadres en een I/O-adres heeft de processor een $M/\overline{IO}$-lijn. Als het geheugen geadresseerd moet worden, maakt de processor deze lijn "H", als een I/O-adres op de adres-bus staat

wordt dit signaal "L" gemaakt. Op deze eenvoudige manier kunnen alle onderdelen van het systeem een duidelijk onderscheid maken tussen de twee soorten adressen die er op de adres-bus kunnen staan. Het systeem reserveert een aantal adressen voor interne doeleinden en men mag deze adressen dus nooit gebruiken voor het aanspreken van uitbreidingen! Omdat oude kaarten, die niet met automatische configuratie zijn uitgerust, vaak met de hand op een bereik in de I/O-adresruimte moeten worden ingesteld, is het dus van het grootste belang te weten welke I/O-adresbereiken gereserveerd zijn en welke vrij gebruikt mogen worden. In de lijst van figuur 3/19.10-2 zijn alle I/O-adresbereiken weergegeven (natuurlijk in hexadecimale notatie) en de functie die er aan is toegekend.

000	-	01F	eerste DMA-controller
020	-	03F	master interrupt-controller
040	-	05F	timer (8254)
060	-	06F	toetsenbord-controller (8042)
070	-	07F	real-time klok
080	-	09F	DMA, page register
0A0	-	0BF	slave interrupt-controller
0C0	-	0DF	tweede DMA-controller
0F0	-	0FF	co-processor
100	-	16F	vrij
170	-	178	tweede IDE-controller
179	-	1CD	grafische kaart VGA/8514
1F0	-	1F8	eerste IDE-controller
1F9	-	1FF	vrij
200	-	207	vrij of spel-poort
210	-	217	uitbreidingseenheid
220	-	23B	vrij, wordt gebruikt door de soundblaster
23C	-	23F	vrij, wordt gebruikt door een bus-muis
250	-	277	vrij
278	-	27F	tweede parallelle poort LPT2
2A0	-	2E7	vrij
2E8	-	2EF	grafische kaart VGA/8514 of COM4
2F8	-	2FF	seriële poort COM2
300	-	31F	vrij, soms gebruikt door een SCSI-adapter
330	-	377	vrij, eventueel in gebruik door MIDI
378	-	37F	eerste parallelle poort LPT1
390	-	39F	vrij
3B0	-	3BF	monochrome grafische kaart
3C0	-	3CF	grafische kaart EGA/VGA
3D0	-	3DF	grafische kaart CGA
3E0	-	3EF	vrij
3E8	-	3EF	vrij, eventueel COM3
3F0	-	3F7	floppy-controller
3F8	-	3FF	seriële poort COM1

Figuur 3/19.10-2: Het I/O-adresbereik van de moderne PC.

19.10 De opstart-procedure van de PC

IRQ0	timer uitgang
IRQ1	toetsenbord
IRQ2	omleiding naar de tweede interrupt-controller voor IRQ8 tot en met IRQ15
IRQ3	seriële poort COM2
IRQ4	seriële poort COM1
IRQ5	parallelle poort LPT2
IRQ6	diskette-controller
IRQ7	parallelle poort LPT1
IRQ8	klok-IC
IRQ9	softwarematige omleiding voor IRQ2 (INT 0Ah)
IRQ10	gereserveerd
IRQ11	gereserveerd
IRQ12	gereserveerd
IRQ13	co-processor
IRQ14	harde schijf controller
IRQ15	SCSI-controller

Figuur 3/19.10-3: De gereserveerde hardware-interrupts.

Interrupt-adressen

Een interrupt is een hardware-signaal van een perifere schakeling naar de processor, waardoor deze onmiddellijk ophoudt met de opdrachten waarmee hij bezig is en beschikbaar komt voor de perifere schakeling die de interrupt heeft aangevraagd. Zonder dit interrupt-mechanisme zou de processor voortdurend allerlei gebeurtenissen moeten aftasten om te controleren of alles volgens plan verloopt.

Soms is de processor met een actie bezig die niet onderbroken mag worden, tenzij in absolute noodgevallen. Vandaar dat het begrip "NMI" of "non maskable interrupt" is ingevoerd. De NMI wordt gebruikt om de processor vrij te maken in absolute noodgevallen, zoals problemen met het geheugen.

Naast de besproken hardware-interrupts kan de processor ook onderbroken worden door zogenoemde software-interrupts. In totaal kan men 256 interrupts (hard- + software) invoeren. Deze worden door de processor op dezelfde manier behandeld als hardware-interrupts. Deze interrupts worden aangeduid met "INT", gevolgd door een nummer, zoals INT 21h.

De onderste 1.024 bytes van het geheugen zijn gereserveerd voor het opslaan van de 256 adressen van de interrupts. Als een interrupt optreedt vermenigvuldigt de processor het interrupt-nummer met vier en gaat naar het betreffende adres in het geheugen. Op dit adres start de routine, die de betreffende interrupt moet behandelen.

Een overzicht van alle gereserveerde hardware-interrupts is gegeven in figuur 3/19.10-3.

Een overzicht van de gereserveerde software-interrupts treft men aan in de lijst die in de figuren 3/19.10-4 en -5 is gegeven.

19.10 De opstart-procedure van de PC

000 - 003	INT0	divide by zero
004 - 007	INT1	single step IRET
008 - 00B	INT2	non maskable interrupt
00C - 00F	INT3	breakpoint
010 - 013	INT4	overflow IRET
014 - 017	INT5	print screen
018 - 01F	INT6	gereserveerd 018-01B en 01C-01F
020 - 023	INT8	time of day ticker IRQ0
024 - 027	INT9	toetsenbord IRQ1
028 - 02B	INTA	IRQ2 omleiding naar IRQ9
02C - 02F	INTB	COM2 communications, IRQ3
030 - 033	INTC	COM1 communications, IRQ4
034 - 037	INTD	parallelle poort 1, IRQ5
038 - 03B	INTE	floppy disk, IRQ6
03C - 03F	INTF	parallelle poort 1, IRQ7, slave 8259
040 - 043	INT10	ROM-handler video
044 - 047	INT11	ROM-handler, equipment check
048 - 04B	INT12	ROM-handler, memory check
04C - 04F	INT13	ROM-handler, diskette I/O
050 - 053	INT14	ROM-handler, COM I/O
054 - 057	INT15	ROM catchall handlers
058 - 05B	INT16	ROM-handler, toetsenbord I/O
05C - 05F	INT17	ROM-handler, printer I/O
060 - 063	INT18	ROM-handler, BASIC startup
064 - 067	INT19	ROM-handler, bootstrap
068 - 06B	INT1A	ROM-handler, time of day
06C - 06F	INT1B	ROM-handler, BREAK toetsenbord
070 - 073	INT1C	ROM-handler, gebruikers-ticker
074 - 077	INT1D	ROM-pointer, video-initialization
078 - 07B	INT1E	ROM-pointer, diskette parameters
07C - 07F	INT1F	ROM-pointer, graph. character Set 2
080 - 083	INT20	DOS - terminate program
084 - 087	INT21	DOS - function call
088 - 08B	INT22	DOS - program's terminate adres
08C - 08F	INT23	DOS - program's control-break Adres
090 - 093	INT24	DOS - critical error handler
094 - 097	INT25	DOS - absolute disk read
098 - 09B	INT26	DOS - absolute disk write
09C - 09F	INT27	DOS - TSR terminate & stay ready
0A0 - 0FF	INT28	DOS - idle loop, IRET
	t/m	
	INT3F	DOS - idle loop, IRET
100 - 103	INT40	hard disk pointer, floppy handler
104 - 107	INT41	ROM-pointer
108 - 10B	INT42	gereserveerd
	t/m	
	INT45	gereserveerd
10C - 10F	INT46	hard disk parameters

Figuur 3/19.10-4: De gereserveerde software-interrupts, deel 1.

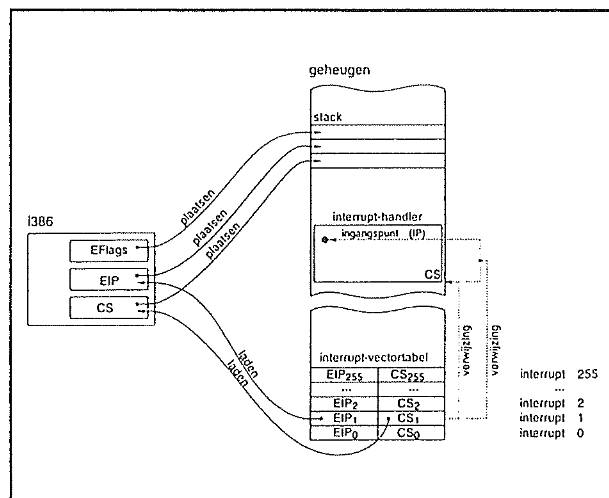
110 - 17F	INT47	gereserveerd
	t/m	
	INT5F	gereserveerd
180 - 19F	INT60	gereserveerd voor gebruikers
	t/m	
	INT67	gereserveerd voor gebruikers (67 is expanded-geheugen)
1A0 - 1BF	INT68	niet gebruikt
	t/m	
	INT6F	niet gebruikt
1C0 - 1C3	INT70	real time clock, IRQ8
1C4 - 1C7	INT71	omleiding naar IRQ2, IRQ9, LAN-adapter 1
1C8 - 1CB	INT72	gereserveerd, IRQ10
1CC - 1CF	INT73	gereserveerd, IRQ11
1D0 - 1D3	INT74	gereserveerd, IRQ12
1D4 - 1D7	INT75	80287 fout naar NMI, IRQ13
1D8 - 1DB	INT76	hard disk, IRQ14
1DC - 1DF	INT77	gereserveerd, IRQ15
1E0 - 1FF	INT78	niet gebruikt
	t/m	
	INT7F	niet gebruikt
200 - 217	INT80	gereserveerd voor BASIC
	t/m	
	INT85	gereserveerd voor BASIC
218 - 21B	INT86	Net-BIOS, relocated interrupt 18H
218 - 3C3	INT87	
	t/m	
	INTF0	gereserveerd voor BASIC-interpreter
3C4 - 3FF	INTF1	niet gebruikt
	t/m	
	INTFF	niet gebruik

Figuur 3/19.10-5: De gereserveerde software-interrupts, deel 2.

Interrupt-vectoren

Iedere software-interrupt heeft een zogenoemde "interrupt-vector" die verwijst naar het geheugen-adres van de "interrupt-handler" die de interrupt afhandelt, zie figuur 3/19.10-6.

Iedere software-interrupt wordt aangemeld in de vector-tabel en wel op de adressen die gegeven zijn in de tabellen van de figuren 3/19.10-7 en -8.



Figuur 3/19.10-6: Het aanroepen van software-interrupts met interrupt-vectoren.

19.10 De opstart-procedure van de PC

0000	01C5:3AD2	divide overflow
0001	01C5:16CF	single step
0002	F000:E2C3	non maskable interrupt
0003	01C5:16CF	breakpoint
0004	F000:FF47	overflow
0005	1701:1A42	ROM-BIOS print screen
0006	E146:FBD0	gereserveerd
0007	F000:FF53	gereserveerd
0008	1701:1BC1	IRQ0 system timer
0009	1701:1F0A	IRQ1 toetsenbord
000A	F000:FF47	IRQ2 gereserveerd
000B	F000:FF47	IRQ3 gereserveerd
000C	F000:FF47	IRQ4 gereserveerd
000D	F000:FF47	IRQ5 gereserveerd
000E	F000:EF57	IRQ6 floppy disk
000F	F000:FF47	IRQ7 gereserveerd
0010	E141:FC20	ROM-BIOS video services
0011	F000:F84D	ROM-BIOS equipment determination
0012	F000:F841	ROM-BIOS memory-size service
0013	1701:1EC8	ROM-BIOS disk services
0014	E137:FCC0	ROM-BIOS communication services
0015	1701:15CD	ROM-BIOS diverse services
0016	1701:1A77	ROM-BIOS toetsenbord services
0017	E111:FF20	ROM-BIOS printer services
0018	F000:F5C4	ROM-BASIC
0019	E144:FBF0	ROM-BIOS bootstrap
001A	F000:FE6E	ROM-BIOS time of day service
001B	01C5:5F4D	ROM-BIOS ctrl-break handler
001C	F000:FF53	ROM-BIOS timer tick
001D	F000:F0A4	ROM-BIOS video parameter table
001E	F000:F13C	ROM-BIOS floppy disk parameters
001F	C000:6937	ROM-BIOS font (characters 80H-FFH)
0020	1701:15EC	MS-DOS process termination
0021	1701:17B7	MS-DOS general services
0022	1951:01C5	MS-DOS terminate adres
0023	1951:0208	MS-DOS ctrl-c handler
0024	1951:0765	MS-DOS critical error handler
0025	1701:1EDE	MS-DOS absolute disk read service
0026	1701:1EF1	MS-DOS absolute disk write service
0027	01C5:4560	MS-DOS TSR-service
0028	1701:1E5D	MS-DOS idle interrupt
0029	01C5:5F42	MS-DOS gereserveerd
002A	F197:0006	MS-DOS gereserveerd
002B	F197:0006	MS-DOS gereserveerd
002C	F197:0006	MS-DOS gereserveerd
002D	F197:0006	MS-DOS gereserveerd
002E	092F:03EB	MS-DOS gereserveerd
002F	0A9B:0969	MS-DOS multiplex interrupt
0030	C515:5FEA	MS-DOS gereserveerd
0031	F197:0001	MS-DOS gereserveerd
0032	F197:0006	MS-DOS gereserveerd
0033	008D:0005	muis-driver
0034	F197:0006	MS-DOS gereserveerd
0035	F197:0006	MS-DOS gereserveerd
0036	F197:0006	MS-DOS gereserveerd
0037	F197:0006	MS-DOS gereserveerd
0038	F197:0006	MS-DOS gereserveerd
0039	F197:0006	MS-DOS gereserveerd
003A	F197:0006	MS-DOS gereserveerd
003B	F197:0006	MS-DOS gereserveerd
003C	F197:0006	MS-DOS gereserveerd
003D	F197:0006	MS-DOS gereserveerd
003E	F197:0006	MS-DOS gereserveerd
003F	F197:0006	MS-DOS gereserveerd
0040	F000:EC59	ROM-BIOS floppy-disk driver

Figuur 3/19.10-7: Adressen van de vector-tabel van de software-interrupts, deel 1.

19.10 De opstart-procedure van de PC

0041	F000:E671	ROM-BIOS fixed disk parameters
0042	F000:F065	ROM-BIOS default video driver (EGA)
0043	C000:6537	EGA, MCGA, VGA character table
0044	F000:FF47	ROM-BIOS font (characters 00H-7FH)
0045	F000:FF47	gereserveerd
0046	F000:E681	ROM-BIOS fixed disk param. (drive 1)
0047	F000:FF47	gereserveerd
0048	F000:FF47	gereserveerd
0049	F000:FF47	gereserveerd
004A	F000:FF47	ROM-BIOS alarm handler
004B	E104:FFF0	gereserveerd
tot en met		
005F	F000:FF47	gereserveerd
0060	E143:FC00	gebruikers-interrupts
0061	0000:0000	gebruikers-interrupts
0062	0000:0000	gebruikers-interrupts
0063	0000:0000	gebruikers-interrupts
0064	0000:0000	gebruikers-interrupts
0065	0000:0000	gebruikers-interrupts
0066	1701:19F6	gebruikers-interrupts
0067	0070:0020	expanded memory driver
0068	F000:FF47	gereserveerd
0069	F000:FF47	gereserveerd
006A	F000:FF47	gereserveerd
006B	F000:FF47	gereserveerd
006C	F000:FF47	gereserveerd
006D	C000:344F	gereserveerd
006E	F000:FF47	gereserveerd
006F	F000:FF47	gereserveerd
0070	F000:F443	IRQ8 CMOS real-time clock
0071	F000:81A7	IRQ9 gereserveerd
0072	F000:FF47	IRQ10 gereserveerd
0073	F000:FF47	IRQ11 gereserveerd
0074	E132:FD10	IRQ12 gereserveerd
0075	F000:9A44	IRQ13 gereserveerd
0076	F000:A1AF	IRQ14 gereserveerd
0077	F000:FF47	IRQ15 gereserveerd
0078	0000:0000	gereserveerd
0079	0000:0000	gereserveerd
007A	0000:0000	gereserveerd
007B	0000:0000	gereserveerd
007C	0000:0000	gereserveerd
007D	0000:0000	gereserveerd
007E	0000:0000	gereserveerd
007F	0000:0000	gereserveerd
0080	0000:0000	gereserveerd voor BASIC
tot en met		
00F0	00D0:0000	gereserveerd voor BASIC
00F1	A1A1:0000	gebruikers-interrupts
00F2	01F0:000A	gebruikers-interrupts
00F3	E671:0001	gebruikers-interrupts
00F4	A180:00E0	gebruikers-interrupts
00F5	A110:8235	gebruikers-interrupts
00F6	0001:A07D	gebruikers-interrupts
00F7	9C90:9D03	gebruikers-interrupts
00F8	0000:0000	gebruikers-interrupts
00F9	7ECF:0005	gebruikers-interrupts
00FA	00F4:7C00	gebruikers-interrupts
00FB	0180:7C00	gebruikers-interrupts
00FC	0050:0001	gebruikers-interrupts
00FD	0000:7F98	gebruikers-interrupts
00FE	7EA0:7296	gebruikers-interrupts
00FF	0080:7FCE	gebruikers-interrupts

Figuur 3/19.10-8: Adressen van de vector-tabel van de software-interrupts, deel 2.

19.10 De opstart-procedure van de PC

De POST foutherkenning

Inleiding

De testroutines van de POST onderzoeken de volledige hardware van de computer en kunnen auditieve of visuele foutcodes genereren. Aan de hand van deze toontjes of codes kan de gebruiker/ster, die iets afweet van de hardwarematige samenstelling van zijn systeem, in vele gevallen de oorzaak van een storing lokaliseren en soms zelfs repareren.

Het is dus van het grootste belang dat iedere gebruiker/ster goed op de hoogte is van de betekenis van deze toontjes en codes! Want ook als men zichzelf niet in staat acht fouten te repareren in een elektronische schakeling, dan kan het vermelden van de foutcodes bij het telefonisch melden van de storing bij de technische dienst van de leverancier vaak helpen om de storing veel efficiënter en goedkoper te verhelpen.

Belangrijke opmerking

Iedereen weet dat er in PC-wereld iedere maand mooie afspraken worden gemaakt, maar dat alle fabrikanten die afspraken weer aan hun laars lappen als het zo uitkomt. Het zal bekend zijn dat er diverse fabrikanten zijn, die zich gespecialiseerd hebben in het ontwerpen van de software die in ROM zit. Deze software wordt algemeen samengevat met de term "BIOS". De POST maakt deel uit van dit BIOS. Een van de bekendste BIOS-leveranciers is AMI, afkorting van "American Megatrends Incorporated". In de meeste goedkope klonen is een BIOS van deze leverancier ingebouwd.

Uiteraard levert ook IBM, die aan de wieg van het PC-gebeuren heeft gestaan, BIOS-versies. Daarnaast is ook het Amerikaanse Phoenix een bekende leverancier van BIOS-software.

Maar behalve verschillende leveranciers zijn er uiteraard ook de nodige BIOS-versies op de markt verschenen! Het probleem is dat niet alle leveranciers en alle versies dezelfde codes voor foutherkenning leveren.

In de navolgende opsommingen worden alleen die codes beschreven, die volledig onafhankelijk ***zouden moeten zijn*** van leverancier of versie.

Men noemt dit de "elementaire foutcodes" en iedere BIOS-fabrikant zou deze, volgens ooit gemaakte afspraak, moeten respecteren. Helaas is de realiteit iets ingewikkelder...

Fouten

In principe kan men fouten indelen in de volgende twee categorieën:

- hardware fouten;
- software fouten.

Hardware fouten zijn fouten die een technisch mankement van de computer aangeven. Voorbeelden hiervan zijn een verkeerde videokaart of een hard disk die niet (correct) aangesloten is.

Software fouten zijn fouten die een logisch mankement van de computer aangeven. Voorbeelden hiervan zijn verkeerde settings in de file CONFIG.SYS of een TSR (Terminate and Stay Resident) programma dat geheugenruimte gebruikt die andere programma's ook nodig hebben.

Foutcodes van de POST

Bij het opstarten van het systeem kunnen er foutcodes op het scherm verschijnen. Deze foutcodes zijn een zeer handig hulpmiddel om een hardware fout in het systeem te lokaliseren.

- Foutmelding 02X:

Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij de controle van de voeding. Oorzaken kunnen bijvoorbeeld zijn dat een zekering van een spanningsgroep uitgevallen is of dat er een voedings-IC defect is.

- Foutmelding 1XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij de test van het moederbord. Voorbeelden hiervan zijn bijvoorbeeld DMA-problemen of problemen met het toetsenbord. Deze groep is nog verder uitgesplitst en staat beschreven in de tabel van figuur 3/19.10-9.
- Foutmelding 20X:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij de test van het geheugen.
  Hierbij moet men in eerste instantie denken aan IC's die niet correct meer functioneren. Bij deze foutcode verschijnt er nog een tweede waarde achter de foutmelding, namelijk de locatie van de foutieve chip.
- Foutmelding 30X:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij de controle van het toetsenbord. Hier zijn de mogelijke oorzaken een toets die vast zit of dat er bij de configuratie een verkeerd toetsenbord opgegeven is. Als er een toets vast zit, dan wordt er nog een extra code meegegeven die aangeeft welke toets vast zit.
- Foutmelding 4XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het controleren van de opstart-instelling van de videokaart. Gedurende de POST wordt zelfs de beste grafische kaart immers nog steeds ingesteld als een monochrome adapter. Mogelijkheden hier zijn dat het geheugen van de kaart niet goed functioneert of dat de configuratie verkeerd ingesteld is. Deze groep is nog verder uitgesplitst en staat volledig beschreven in de tabel van figuur 3/19.10-10.

Code	Fout
101	Main system board failed
102	Main system board failed
103	Main system board failed
104	Main system board failed
105	Keyboard connector error
106	Main system board failed
107	Option adapter failure
108	Main system board failed
109	Direct memory access test error
121	Unexpected hardware interrupts occured
131	Cassette wrap test failure
152	Main system board failed
199	User indicated configuration not correct

Figuur 3/19.10-9: Overzicht van de foutmeldingen van het moederbord.

Code	Fout
401	Monochrome memory test, horizontal sync frequency test or video test failed
408	User indicated display attributes failure
416	User indicated character set failure
424	User indicated 80X25 mode failure
432	Parallel port test failed (monochrome adapter)

Figuur 3/19.10-10: De foutcodes bij de test van de monochrome modes van de grafische kaart.

19.10 De opstart-procedure van de PC

Code	Fout
601	Diskette power on diagnostics test failed
602	Diskette test failed
606	Diskette verify function failed
607	Write protected diskette
608	Bad command diskette status returned
610	Diskette initialization failed
611	Timeout - diskette status returned
612	Bad NEC - diskette status returned
613	Bad DMA - diskette status returned
621	Bad seek - diskette status returned
622	Bad Crc - diskette status returned
623	Record not found - diskette status returned
624	Bad address mark - diskette status returned
625	Bad NEC seek - diskette status returned
626	Diskette data compare error

Figuur 3/19.10-11: De foutcodes bij het testen van de floppy-drives.

- Foutmelding 5XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het controleren van de CGA-modes van de videokaart.
  Ook hier zijn de meest voorkomende oorzaken een niet goed werkende monitor of een verkeerd ingestelde configuratie.
- Foutmelding 6XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij de controle van de floppy disk drive. Hier zou men bijvoorbeeld kunnen denken aan fouten die bij de diagnose test naar voren komen of een time-out van de controller. Deze groep is nog verder uitgesplitst en staat als dusdanig volledig beschreven in de tabel van figuur 3/19.10-11.
- Foutmelding 7XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van de coprocessor.
- Foutmelding 9XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij de test van de parallelle poort. Hier bestaan de mogelijkheden uit weinig meer dan een foutieve initialisatie of een defecte kaart.
- Foutmelding 11XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van de seriële kaart. Mogelijke oorzaken zijn een verkeerde aansluiting of initialisatie.
- Foutmelding 12XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van een alternatieve asynchrone kaart.

19.10 De opstart-procedure van de PC

Code	Fout
1510	8255 Port B failure
1511	8255 Port A failure
1512	8255 Port C failure
1513	8253 Timer 1 did not reach terminal count
1514	8253 Timer 1 stuck on
1515	8253 Timer 0 did not reach terminal count
1516	8253 Timer 0 stuck on
1517	8253 Timer 2 did not reach terminal count
1518	8253 Timer 2 stuck on
1519	8273 Port B error
1520	8273 Port A error
1521	8273 command/read timeout
1522	Interrupt level 4 failure
1523	Ring indicate stuck on
1524	Receive clock stuck on
1525	Transmit clock stuck on
1526	Test indicate stuck on
1527	Ring indicate not on
1528	Receive clock not on
1529	Transmit clock not on
1530	Test indicate not on
1531	Data set ready not on
1532	Carrier detect not on
1533	Clear to send not on
1534	Data set ready stuck on
1536	Clear to send stuck on
1537	Level 3 interrupt failure
1538	Receive interrupt results error
1539	Wrap data miscompare
1540	DMA channel 1 error
1541	DMA channel 1 error
1542	Error in 8273 error checking or status reporting
1547	Stray interrupt level 4
1548	Stray interrupt level 3
1549	Interrupt presentation sequence timeout

Figuur 3/19.10-12: De foutcodes bij het testen van de "Synchronous Data Link Control".

Hier zijn dezelfde mogelijkheden beschikbaar als bij de vorige groep, namelijk verkeerde aansluitingen of foutieve configuratie instellingen.

- Foutmelding 13XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij de test van de spelkaart. Een voor de hand liggende oorzaak is een joystick aansluiting die niet correct functioneert.
- Foutmelding 14XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van de printer. Mogelijke fouten zijn printers die niet "klaar" staan of aansluitingen die niet correct zijn.

19.10 De opstart-procedure van de PC

- Foutmelding 15XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van de SDLC (Synchronous Data Link Control) communicatie kaart. Mogelijke fouten zijn een signaal dat niet aanwezig is of een time-out van een IC. Deze groep is nog verder uitgesplitst en staat beschreven in de tabel van figuur 3/19.10-12.
- Foutmelding 17XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van de harde schijf. Fouten die hier kunnen optreden zijn bijvoorbeeld een controller die niet (goed) reageert of leesfouten van een hard disk. Deze groep is nog verder uitgesplitst en staat beschreven in de tabel van figuur 3/19.10-13.
- Foutmelding 18XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van een van de uitbreidingskaarten. Fouten die hierbij vaak optreden zijn signalen die niet correct overkomen of het niet kunnen disablen/enablen van de verschillende testen. Deze groep is nog verder uitgesplitst en staat beschreven in de tabel van figuur 3/19.10-14.

Code	Fout
1701	Fixed disk post error
1702	Fixed disk adapter error
1703	Fixed disk drive error
1704	Fixed disk adapter or drive error
1780	HDD 0 Reset failure
1781	HDD 1 Reset failure
1782	Controller reset failure
1790	HDD 0 Read failure
1791	HDD 1 Read failure

Figuur 3/19.10-13: De foutcodes van het testen van de harde schijf controller.

Code	Fout
1801	I/O Expansion unit Post error
1810	Enable/Disable failure
1811	Extended card wrap test failed (disabled)
1812	High order address lines failure (disabled)
1813	Wait state failure (disabled)
1814	Enable/Disable could not be set on
1815	Wait state failure (enabled)
1816	Extended card wrap test failed (enabled)
1817	High order address lines failure (enabled)
1818	Disable not functioning
1819	Wait request switch not set correctly
1820	Receiver card wrap test failure
1821	Receiver high order address lines failure

Figuur 3/19.10-14: Foutcodes bij het testen van uitbreidingskaarten.

19.10 De opstart-procedure van de PC

Code	Fout
2010	8253 Port A failure
2011	8253 Port B failure
2012	8253 Port C failure
2013	8253 Timer 1 did not reach terminal count / stuck
2014	8253 Timer 1 stuck on
2016	8253 Timer 2 did not reach terminal count / stuck
2017	8251 Data set ready failed to come on
2018	8251 Clear to send not sensed
2019	8251 Data set ready stuck on
2020	8251 Clear to send stuck on
2021	8251 Hardware reset failed
2022	8251 Software reset failed
2023	8251 Software "error reset" failed
2024	8251 Transmit ready did not come on
2025	8251 Receive ready did not come on
2026	8251 could not force "overrun" error status
2027	Interrupt failure-no timer interrupt
2028	Interrupt failure-transmit, replace card or planar
2029	Interrupt failure-transmit, replace card
2030	Interrupt failure-receive, replace card or planar
2031	Interrupt failure-receive, replace card
2033	Ring indicate stuck on
2034	Receive clock stuck on
2035	Transmit clock stuck on
2036	Test indicate stuck on
2037	Ring indicate not on
2038	Receive clock not on
2039	Transmit clock not on
2040	Test indicate not on
2041	Data set ready not on
2042	Carrier detect not on
2043	Clear to send not on
2044	Data set ready stuck on
2045	Carrier detect stuck on
2046	Clear to send stuck on
2047	Unexpected transmit interrupt
2048	Unexpected receive interrupt
2049	Transmit data did not equal receive data
2050	8251 detected overrun error
2051	Lost data set ready during data wrap
2052	Receive timeout during data wrap

Figuur 3/19.10-15: Foutcodes bij het testen van een "Binary Synchronous Control"-kaart.

- Foutmelding 20XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van de BSC (Binary Synchronous Control) kaart. Fouten die hier mogelijk zijn, zijn bijvoorbeeld de time-out van een signaal of signalen die niet (goed) aanwezig zijn. Deze groep is nog

19.10 De opstart-procedure van de PC

verder uitgesplitst en de subcodes staat beschreven in de tabel van figuur 3/19.10-15.

- Foutmelding 21XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van een alternatieve BSC kaart. Hier gelden dezelfde fouten als bij de vorige beschrijving, namelijk een signaal dat niet (goed) aanwezig is of signalen waar een time-out op gegeven wordt.
- Foutmelding 22XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van de cluster adapter. Fouten die hier aanwezig zijn, zijn onder andere een verkeerd signaal of een defecte kaart.
- Foutmelding 24XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van de EGA-modes van de grafische kaart. Fouten die hier tevoorschijn kunnen komen zijn bijvoorbeeld een foutief geheugendeel op de kaart of een monitor die niet goed functioneert.
- Foutmelding 30XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van de netwerk kaart. Hier kunnen verschillende fouten optreden, zoals verkeerde initialisatiewaarden of signalen die niet goed verschijnen.
- Foutmelding 31XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van de alternatieve netwerk kaart. Hier gelden dezelfde fouten als bij de vorige test, namelijk foutieve initialisatie waarden of signalen die niet aanwezig zijn.
- Foutmelding 36XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van de eventueel aanwezige GPIB (General Purpose Interface Bus). Deze bus wordt gebruikt bij de IEEE488 (Institute of Electrical and Electronic Engineers) standaardbus. Fouten die hier voorkomen zijn onder andere verkeerde aansluitingen of foutieve signaleringen.
- Foutmelding 38XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van een eventueel aanwezige data acquisition kaart. Deze kaart wordt gebruikt voor het verzamelen en registreren van externe analoge gegevens, zoals elektrische signalen van temperatuuropnemers en druksensoren. Deze gegevens zijn afkomstig van externe fysieke eenheden. Fouten die hier tevoorschijn zullen komen zijn vaak foutieve initialisatie waarden of IC's die niet meer goed functioneren.
- Foutmelding 39XX:
  Deze groep van foutmeldingen geeft aan dat er een fout opgetreden is bij het testen van de VGA-modes van de aanwezige grafische kaart. Hier kunnen hetzelfde soort fouten opgespoord worden die ook bij andere videokaarten naar voren kunnen komen, namelijk een foutieve configuratie of een monitor die niet werkt.

Besluit

Het zal duidelijk zijn dat de foutcodes die door de POST gegenereerd worden een zeer nuttig hulpmiddel zijn bij het lokaliseren van de exacte plaats van de fout! Dank zij deze errorcodes kan men niet-fatale fouten, die niet het gehele systeem lamleggen vaak zonder deskundige hulp opsporen en verhelpen. Want vele fouten

ontstaan bijvoorbeeld door het verslepen van en stoten tegen systeemkasten. De mechanische belastingen die daarvan het gevolg zijn kunnen stekers in hun connectoren en IC's in hun voetjes lostrillen, waardoor slechte elektrische contacten ontstaan of zelfs signaalonderbrekingen.

De audio foutcodes van de POST
Minder bekend is dat de meeste BIOS-fabrikanten ook de mogelijkheid hebben ingebouwd om auditieve foutcodes te genereren. Zelfs als de POST zeer fundamentele hardwarefouten detecteert, waardoor het systeem niet meer kan opstarten, bestaat de mogelijkheid dat de elektronica die de ingebouwde luidspreker stuurt het nog wél doet. Dank zij deze auditieve meldingen is men dan toch in staat de plaats van de hardwarefout te vinden. De audiocodes bestaan uit pieptoontjes met een constante frequentie, die ongeveer 0,25 s duren. De toontjes kunnen afzonderlijk optreden met korte tussenpauzes, maar ook gegroepeerd met langere pauzes tussen de groepen.

Voorbeelden:
- ●●●●●
  vijf afzonderlijke piepjes
- ●● ●● ●●●●
  drie groepen van respectievelijk twee, twee en vier piepjes, waarbij iedere groep gescheiden wordt door een langere pauze.

Beschrijving
Hieronder volgt een opsomming van de audio-codes die min of meer gestandaardiseerd zijn.
- ●
  Fout bij het verversen van de dynamische RAM-IC's van het systeemgeheugen.
- ●●
  Fout in de elektronica die de pariteit van de geheugen-inhoud controleert.
- ●●●
  Fout in de onderste 64 kB van het systeem-geheugen.
- ●●●●
  Fout in de timer-elektronica.
- ●●●●●
  Fataal defect in de hoofdprocessor.
- ●●●●●●
  Fout in de belangrijke adreslijn A20.
- ●●●●●●●
  Fout in de virtual mode van de processor.
- ●●●●●●●●
  Defect in het video-geheugen.
- ●●●●●●●●●
  Fout in het ROM-geheugen van het BIOS, de proefsom die bij de zelftest wordt gemaakt om de integriteit van de code te controleren klopt niet.
- ● ● ●●●
  Fout in de altijd onder voedingsspanning blijvende CMOS geheugen-IC's die de opstartcondities van de hardware bevatten en die door de SET-UP van het systeem worden beschreven.
- ● ● ●●●●
  Fout in de geheugen-IC's waarin het BIOS zit.
- ● ●● ●
  Fout in de elektronica, die de timerfuncties van het systeem verzorgt.
- ● ●● ●●
  Fout in de DMA-controller, de schakeling die de directe toegang tot het geheugen regelt voor perifere schakelingen.
- ● ●● ●●●
  Fout in het pagina-register van de DMA-controller.
- ● ●●● ●

19.10 De opstart-procedure van de PC

Fout bij de schakelingen die het proces verzorgen voor het verversen van de informatie in de dynamische geheugen-IC's, de zogenoemde refresh.
- ● ●●● ●●●
  Fout in de eerste 64 kB van het systeemgeheugen.
- ● ●●● ●●●●
  Fout in de eerste 64 kB van het systeemgeheugen.
- ● ●●●● ●
  Fout in de adresseer-logica van de eerste 64 kB van het systeem-geheugen.
- ● ●●●● ●●
  Fout in de pariteitscontrole van de eerste 64 kB van het systeem-geheugen.
- ● ●●●● ●●●
  Fout in de timing van de EISA-bus.
- ● ●●●● ●●●●
  Fout bij het testen van niet-maskeerbare interrupts.
- ●● ● ●
  Fout in bit 0 van het RAM-geheugen.
- ●● ● ●●
  Fout in bit 1 van het RAM-geheugen.
- ●● ● ●●●
  Fout in bit 2 van het RAM-geheugen.
- ●● ● ●●●●
  Fout in bit 3 van het RAM-geheugen.
- ●● ●● ●
  Fout in bit 4 van het RAM-geheugen.
- ●● ●● ●●
  Fout in bit 5 van het RAM-geheugen.
- ●● ●● ●●●
  Fout in bit 6 van het RAM-geheugen.
- ●● ●● ●●●●
  Fout in bit 7 van het RAM-geheugen.
- ●● ●●● ●
  Fout in bit 8 van het RAM-geheugen.
- ●● ●●● ●●
  Fout in bit 9 van het RAM-geheugen.
- ●● ●●● ●●●
  Fout in bit 10 van het RAM-geheugen.
- ●● ●●● ●●●●
  Fout in bit 11 van het RAM-geheugen.
- ●● ●●●● ●
  Fout in bit 12 van het RAM-geheugen.
- ●● ●●●● ●●
  Fout in bit 13 van het RAM-geheugen.
- ●● ●●●● ●●●
  Fout in bit 14 van het RAM-geheugen.
- ●● ●●●● ●●●●
  Fout in bit 15 van het RAM-geheugen.
- ●●● ● ●
  Fout ontdekt bij de test van de DMA-controller, de schakeling die de rechtstreekse toegang tot het geheugen regelt voor de perifere schakelingen.
- ●●● ● ●●
  Fout in de master van de DMA-controller.
- ●●● ● ●●●
  Fout in de master-schakeling van de interrupt-controller.
- ●●● ● ●●●●
  Fout in de slave-schakeling van de interrupt-controller.
- ●●● ●● ●●●●
  Fout ontdekt in de controller van het toetsenbord.
- ●●● ●●● ●●●●
  Fout ontdekt in het geheugensegment, waarin de inhoud van het beeldscherm wordt bewaard.
- ●●● ●●●● ●
  Fout ontdekt in video-RAM.
- ●●● ●●●● ●●
  Fout in het video-BIOS.
- ●●● ●●●● ●●●
  Fout in het video-BIOS ontdekt (video-ROM).
- ●●●● ●● ●
  Fout ontdekt in de timer-chip.
- ●●●● ●● ●●●
  Een fout ontdekt in de belangrijke adreslijn A20.
- ●●●● ●● ●●●●

Niet te verklaren interrupt gedetecteerd.

- ●●●● ●●● ●
  Fout ontdekt in het RAM-geheugen boven de 64 kB grens.
- ●●●● ●●● ●●●●
  Fout ontdekt in de datum- en tijd-chip.
- ●●●● ●●●● ●
  Fout ontdekt in de elektronica die de seriële poort aanstuurt.
- ●●●● ●●●● ●●
  Fout ontdekt in de elektronica die de parallelle poort aanstuurt.
- ●●●● ●●●● ●●●
  Fout ontdekt in de co-processor.

Het BIOS SETUP programma

Configuratiegegevens
PC's beschikken sinds de AT-klasse (PC met 80286-processor) over een zogenoemd CMOS-geheugen waarin de configuratiegegevens van het systeem kunnen worden opgeslagen. Met het bijbehorende SETUP programma kunnen deze instellingen voor het BIOS (Basic Input Output System) worden aangepast om daarmee de prestaties van de PC te optimaliseren. De instellingen betreffen onder meer technische zaken als de klokfrequentie van de bus, waitstates, shadowfuncties en nog veel meer.

AMI BIOS
Het BIOS van de firma AMI (American Megatrends Inc.) is, vooral door de verkoop van PC-klonen, een van de meest verbreide BIOS-varianten. Goede redenen daarvoor zijn het inmiddels redelijk gestandaardiseerde user-interface en de veelzijdige instelmogelijkheden die de SET-UP biedt. In dit afsluitende deel van dit hoofdstuk wordt, aan de hand van het AMI BIOS, aangetoond hoe men door geschikte SETUP-instellingen de prestaties van een PC merkbaar kan opvoeren. Allereerst moet men aan de hand van diverse tests nagaan in hoeverre de SET-UP-instellingen überhaupt in staat zijn de prestatiekenmerken van het systeem, positief of negatief, te beïnvloeden. Vervolgens wordt de AMI BIOS SETUP gedetailleerd besproken om de gebruiker/ster in staat te stellen de afzonderlijke opties zo optimaal mogelijk op de hardware af te stemmen.

BIOS SETUP
Sinds de komst van de AT-PC's, uitgerust met de 80286-processor, zijn begrippen als CMOS of BIOS SETUP vast bestanddeel van de PC-terminologie. In het CMOS-geheugen worden onder meer de configuratiegegevens (informatie over de hardware componenten) van de PC via het SETUP programma ingesteld en opgeslagen. Een batterij of accu zorgt ervoor dat de inhoud van het CMOS-RAM bewaard blijft ook als men de PC uitschakelt. Met de invoering van de NEAT (New Enhanced Advanced Technology) krijgen de begrippen CMOS en BIOS SETUP instellingen nieuwe inhoud. Met behulp van speciale BIOS-programmatuur kunnen de op deze borden aanwezige chipset's hardware-gericht worden ingesteld. Waren de veelzijdige instelmogelijkheden van de NEAT borden bij hun introductie nog een veelbesproken sensatie, tegenwoordig zijn de chipset's en BIOS bouwstenen met vergaande instelmogelijkheden een vanzelfsprekendheid. In elke BIOS bouwsteen van een moderne PC is thans een SETUP-programma aanwezig

19.10 De opstart-procedure van de PC

waarmee de hardware configuratie kan worden opgegeven en de registers van de chipset's gewijzigd kunnen worden. De chipset's op het moederbord en de SET-UP-programma's zijn van producent tot producent verschillend. Onafhankelijk van de gebruikte BIOS-variant zijn daarbij toch altijd de configuratie van de hardware en de instellingen zoals onder meer de klokfrequentie, wait states, shadow functies, cache geheugen, etcetera instelbaar. Hierbij geldt de regel: hoe recenter het BIOS en de chipset, hoe uitgebreider de instelmogelijkheden.

Instelmogelijkheden

De moderne chipset's bieden de gebruiker/ster de mogelijkheid bepaalde registerinstellingen te wijzigen en op deze wijze het systeem op de aanwezige hardwarecomponenten optimaal af te stemmen, met het doel de prestaties van het systeem te verhogen.

Afhankelijk van het BIOS biedt het geïntegreerde SETUP-programma meer of minder instelmogelijkheden. Externe SET-UP-programma's op diskette behoren tot het verleden. De BIOS SETUP is thans vast bestanddeel van de BIOS zélf. Los van het aantal instelmogelijkheden dat het systeem biedt, zijn de aanwezige opties qua uitwerking, tenminste in principe, voor alle beschikbare chipset's min of meer vergelijkbaar.

Iedere SETUP-programma biedt in ieder geval de onderstaande instellingen:

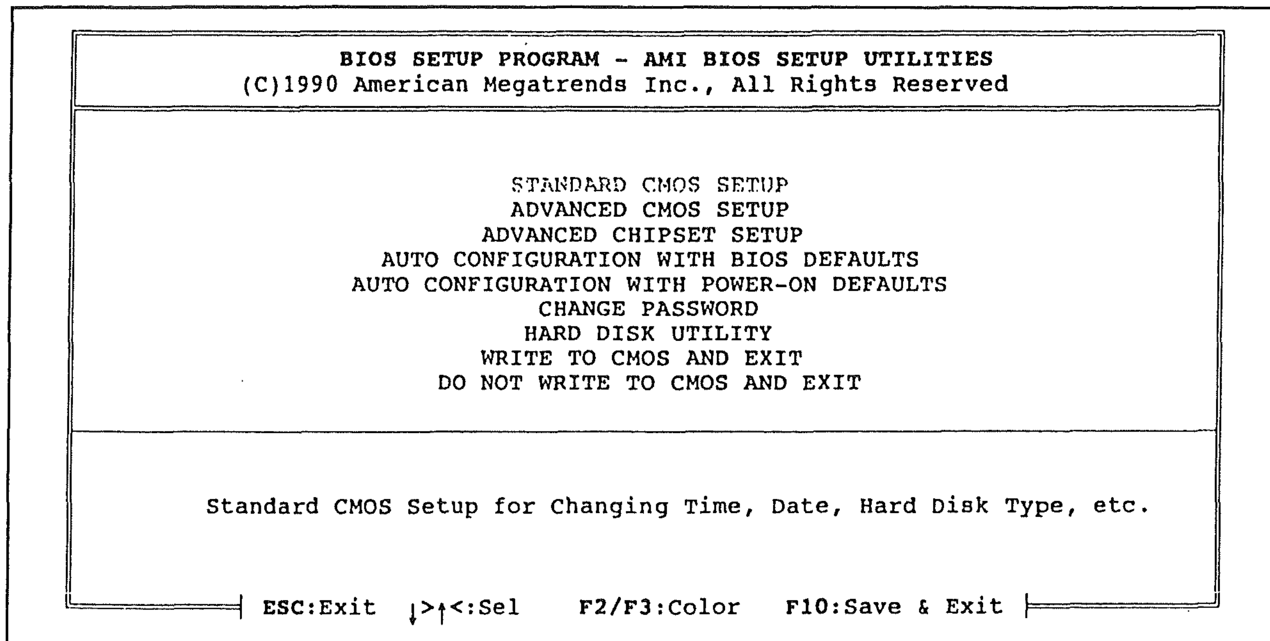
- Tot de belangrijkste instellingen die men kan wijzigen, behoren de wijzigingen van de klokfrequentie voor de verschillende onderdelen van het systeem. Afhankelijk van het SETUP-programma hebben deze vooral betrekking op de klokfrequentie van de bus (en daarmee van de uitbreidingskaarten). Bij sommige SETUP's kan men ook nog andere klokfrequenties instellen, zoals deze van de CPU of de DMA-controller.
- De volgende belangrijke instelling is de vastlegging van de geheugen toegang; wachtcycli bij besturingscommando's en waitstates bij algemene toegang kunnen individueel worden geregeld. Moderne SETUP programma's voor snelle moederborden maken ook nog onderscheid tussen waitstates voor het lezen en schrijven in het interne geheugen; daarbij bestaat in beide gevallen de keuze tussen 0 tot en met 4 waitstates. Deze instellingen zijn van wezenlijk belang voor het prestatie vermogen (de performance) van het systeem.
- Instellingen zoals de grootte en locatie van het EMS alsmede shadow-RAM functies voor ROM BIOS en videokaart behoren eveneens tot de mogelijkheden.
- Indien het moederbord over cachegeheugen beschikt, dan zijn ook hiervoor diverse instelmogelijkheden.

De SETUP van het AMI BIOS

Vooral met de compatibele PC-klonen heeft het moderne AMI BIOS, dat sinds begin 1991 op de markt is, enorme verspreiding gevonden.

Afgezien van aanpassingen aan verder ontwikkelde of nieuwe chipset's zijn het verschijningsbeeld en de bediening van dit BIOS programma sindsdien ongewijzigd gebleven en tegenwoordig bij zeer veel PC's aan te treffen.

De SETUP wordt vlak na het inschakelen van de PC opgeroepen door op de "DEL"- of de "ESC"-toets te drukken. Er verschijnt dan in eerste instantie het hoofdmenu van figuur 3/19.10-16. Daarin kan men kiezen tussen de hoofdtaken van het SET-UP-programma.



Figuur 3/19.10-16: Het hoofdmenu van de AMI BIOS SETUP.

Hoofdmenu

Naast de basisinstellingen van de systeemconfiguratie (STANDARD CMOS SETUP) kunnen van daaruit de geavanceerde CMOS SETUP (ADVANCED CMOS SETUP) en de geavanceerde SETUP van de chipset (ADVANCED CHIPSET SETUP) onder handen worden genomen. Verdere hoofdmenu opties zijn de automatische configuratie (AUTO CONFIGURATION WITH BIOS DEFAULTS en AUTO CONFIGURATION WITH POWER-ON DEFAULTS), de wachtwoord beveiliging (CHANGE PASSWORD) en een hulpprogramma voor de vaste schijf (HARD DISK UTILITY).

Instellingen opslaan

Bij het verlaten van de SETUP kan men in het hoofdmenu bepalen of de ingevoerde wijzigingen in het CMOS-geheugen bewaard zullen worden (WRITE TO CMOS AND EXIT) of niet (DO NOT WRITE TO CMOS AND EXIT). Men bepaalt de keuze in het hoofdmenu met behulp van de cursortoetsen en de "Enter"-toets ter bevestiging.

Standaard CMOS SETUP

Via de standaard CMOS SETUP wordt de configuratie van de PC ingesteld, zie figuur 3/19.10-17. De gegevens moeten bij de eerste in gebruik name van de PC worden ingevoerd. Zonder deze gegevens kan de apparatuur niet correct functioneren. Over het algemeen worden de juiste gegevens wel reeds door de PC dealer ingevoerd.
Ten gevolge van elektromagnetische inductie, een defecte of lege accu of bij de inbouw van een nieuwe of andere hardware-uitbreiding kan het echter zijn dat men de standaard CMOS SETUP zelf moet wijzigen. Men gaat daarvoor met behulp van de cursortoetsen naar de gewenste instelling.

19.10 De opstart-procedure van de PC

```
                BIOS SETUP PROGRAM - STANDARD CMOS SETUP
        (C)1990 American Megatrends Inc., All Rights Reserved

Date (mn/date/year): Fri, Sept 16 1994            Base memory : 640 KB
Time (hour/min/sec): 18 : 59 : 23                 Ext. memory : 3072 KB
Daylight saving    : Disabled        Cyln  Head  WPcom LZone Sect  Size
Hard disk C: type  : 47 = USER TYPE  1024  12    65535 1024  17    102 M
Hard disk D: type  : 47 = USER TYPE  1010  12    65535 1010  55    325 M
Floppy drive A:    : 1.2  MB, 5¼"
Floppy drive B:    : 1.44 MB, 3½"               Sun Mon Tue Wed Thu Fri Sat
Primary display    : VGA/PGA/EGA
Keyboard           : Installed                   30  31   1   2   3   4   5
                                                  6   7   8   9  10  11  12
                                                 13  14  15  16  17  18  19
  Month : Jan, Feb,.....Dec                      20  21  22  23  24  25  26
  Date  : 01, 02, 03,...31                       27  28  29  30   1   2   3
  Year  : 1901, 1902,...2099                      4   5   6   7   8   9  10

  ESC:Exit ↓>↑<:Sel F2/F3:Color PU/PD:Modify
```

Figuur 3/19.10-17: De standaard CMOS SETUP.

Met de toetsen “PageUp” of “PageDown” kan men de instelling dan wijzigen. Deze vorm van besturing geldt trouwens voor alle menu’s van het AMI BIOS SETUP programma.
De standaard SETUP voorziet zowel in de instelling van de zogenoemde systeemdatum en tijd als in de opgaaf van het type vaste schijf (of schijven) en diskettestation(s). Daarnaast wordt hier ook het type van de grafische kaart vastgesteld. Met een druk op de “Esc”-toets keert men uit de standaard CMOS SETUP terug naar het hoofdmenu.

Geavanceerde CMOS SETUP
Het beeldscherm van de geavanceerde CMOS SETUP zal er, afhankelijk van het type PC, in grote lijnen zo uitzien als in figuur 3/19.10-18 weergegeven. Zoals de naamgeving al doet vermoeden, kan men daarin verdergaande wijzigingen in het systeem aanbrengen.
Een korte beschrijving van de betekenis van de verschillende instellingen:

- **Typematic Rate Programming**
  Als de optie “Typematic Rate Programming” is geactiveerd (Enabled), houdt het systeem rekening met de volgende toetsen hoofdinstellingen: “Typematic Rate Delay (msec)” en “Typematic Rate (Chars/sec)”.
- **Typematic rate delay (msec)**
  Met deze optie bepaalt men hoe lang een toets op het toetsenbord ingedrukt moet blijven voordat dit als een herhaling van de toetsaanslag geldt. Deze instelling is alleen werkzaam als de optie “Typematic Rate Programming” is geactiveerd.
- **Typematic rate (chars/sec)**
  Deze optie bepaalt hoe vaak een toetsaanslag wordt herhaald als de toets één seconde lang ingedrukt blijft. De ingestelde waarde geeft dus het aantal tekens per seconde aan.
  De instelling is alleen werkzaam als de optie “Typematic Rate Programming” is geactiveerd.
- **Above 1 MB memory test**

19.10 De opstart-procedure van de PC

```
                BIOS SETUP PROGRAM - ADVANCED CMOS SETUP
         (C)1990 American Megatrends Inc., All Rights Reserved

 Typematic Rate Programming : Disabled   Adaptor ROM Shadow C800,16K: Disabled
 Typematic Rate Delay (msec): 500        Adaptor ROM Shadow CC00,16K: Disabled
 Typematic Rate (Chars/Sec) : 15.0       Adaptor ROM Shadow D000,16K: Disabled
 Above 1 MB Memory Test     : Disabled   Adaptor ROM Shadow D400,16K: Disabled
 Memory Test Tick Sound     : Enabled    Adaptor ROM Shadow D800,16K: Disabled
 Memory Parity Error Check  : Enabled    Adaptor ROM Shadow DC00,16K: Disabled
 Hit <ESC> Message Display  : Enabled    Adaptor ROM Shadow E000,16K: Disabled
 Hard Disk Type 47 Data Area: 0:300      Adaptor ROM Shadow E400,16K: Disabled
 Wait For <F1> If Any Error : Disabled   Adaptor ROM Shadow E800,16K: Disabled
 System Boot Up Num Lock    : Off        Adaptor ROM Shadow EC00,16K: Disabled
 Weitek Processor           : Absent     System  ROM Shadow F000,64K: Enabled
 Floppy Drive Seek At Boot  : Enabled
 System Boot Up Sequence    : C:,A:
 System Boot Up Speed       : High
 Cache Memory               : Enabled
 Password Checking Option   : Disabled
 Video   ROM Shadow C000,16K: Enabled
 Video   ROM Shadow C400,16K: Enabled

     ESC:Exit  ↓>↑<:Sel  (Ctrl)Pu/Pd:Modify  F1:Help  F2/F3:Color
     F5:Old Values  F6:BIOS Setup Defaults   F7:Power-On Defaults
```

Figuur 3/19.10-18: De geavanceerde CMOS SETUP.

De instelling geeft aan of bij het starten van de PC het intern geheugen boven 1 MB al dan niet moet worden gecontroleerd.

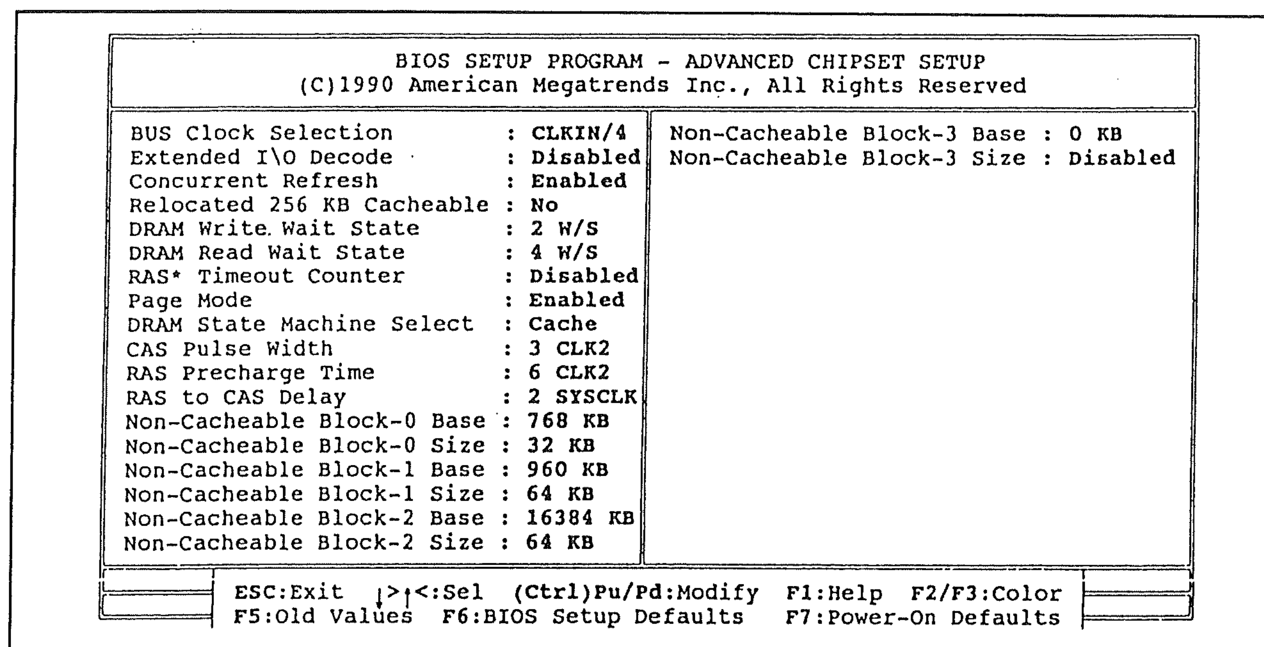
- **Memory test tick sound**
  Hier kan men aangeven of de controle van het geheugen bij het starten van de PC door een tikgeluid mag worden begeleid of niet.
- **Memory parity error check**
  De optie voorziet in het activeren van een pariteitscontrole op de betrouwbaarheid van het aanwezige interne geheugen. Het is ten zeerste aan te bevelen deze controle in te schakelen! Een pariteitsfout duidt er immers meestal op dat er iets mis is met de geheugen-chip's, bijvoorbeeld een iets te lange toegangstijd waardoor men er niet zeker van kan zijn dat de in het geheugen geschreven gegevens betrouwbaar zijn.
- **Hit <ESC> message display**
  Met deze optie kan tijdens het starten van de PC de boodschap worden uitgeschakeld die aangeeft hoe men het SET-UP programma kan oproepen.
- **Hard disk type 47 area**
  Hier kan het te gebruiken geheugengebied worden opgegeven voor zelf gedefinieerde typen vaste schijven van het type 47.
- **Wait for <F1> if any error**
  De optie bepaalt of bij het optreden van een storing of fout tijdens de startfase van de PC een melding verschijnt die om een druk op de “F1”-toets vraagt.
- **System boot up num lock**
  Hiermee kan men instellen of de “NumLock”-toets na het opstarten van de PC in of uitgeschakeld zal zijn. De instelling beïnvloedt de werking van het numerieke toetsenbord als cursorbesturing toetsen of als cijfertoetsen.
- **Weitek processor**
  Als de PC over een wiskundige processor van het merk Weitek beschikt, moet men dit hier opgeven door “Present” in te stellen. Deze instelling is alleen van belang voor bezitters van een nu vol-

strekt verouderde PC van het type 80386 of 80486SX.

- **Floppy drive seek at boot**
  Als men deze optie activeert, dan test het BIOS bij het starten van de PC op de aanwezigheid van diskettestations. Bij de instelling "Disabled" gebeurt dit niet hetgeen de startprocedure enigszins bespoedigt.
- **System boot up sequence**
  Deze optie bepaalt de volgorde waarin het BIOS bij het starten van de PC op de aanwezige schijven zoekt naar het besturingssysteem om dit vervolgens te laden. Men kan daarbij kiezen tussen de volgorde "A:, C:" of "C:, A:". Omdat tegenwoordig zowat elke PC over een vaste schijf beschikt, lijkt het voor de hand te liggen om het zoeken bij "C:" te beginnen. Het besturingssysteem zal daarop bijna altijd worden gevonden zodat niet meer op het diskettestation "A:" moet worden gezocht. Het opstarten verloopt daardoor natuurlijk sneller. Deze optie heeft echter één belangrijk nadeel: ***men kan onder extreme omstandigheden de PC niet meer opstarten van een SYS-diskette als het besturingssysteem op de vaste schijf om de een of andere reden niet meer toegankelijk is!***
- **System boot up speed**
  Veel PC's bieden de mogelijkheid te kiezen tussen twee processorsnelheden (bijvoorbeeld 66 of 133 MHz). Met deze optie kan men aangeven of de hoge of lage klokfrequentie van de centrale processor na het starten van de PC ingesteld zal zijn.
- **Cache memory**
  Afhankelijk van de BIOS-versie zijn hier verschillende waarden mogelijk. Oudere versies bieden alleen de keuze tussen "Enabled" en "Disabled", bij BIOS versies voor 80486 PC's en hoger kan men daarentegen kiezen tussen "Disabled", "Internal" en "Both". Bedoeld wordt in beide gevallen het inschakelen of uitschakelen van het interne of externe processor cache-geheugen. De enige zinvolle keuzes zijn "Enabled" in het ene en "Both" in het andere geval omdat alleen daarmee de processor zijn volle snelheid kan ontplooien.
- **Fast gate 20 option**
  Deze optie leidt het beheer van de zogenaamde A20 adreslijn om, die nodig is voor het adresseren van intern geheugen boven 1 MB. Normaliter wordt deze functie via de toetsenbord controller uitgevoerd. Dat werkt echter aanzienlijk trager dan via de methode die met deze optie kan worden ingeschakeld. De keuze is trouwens maar in een beperkt aantal versies beschikbaar.
- **Password checking option**
  Hiermee kan men de toegang tot de PC door middel van een wachtwoord beveiligen. Men bepaalt hier alleen maar of en wanneer de wachtwoord beveiliging actief zal zijn (voor de invoer of het wijzigen van het wachtwoord zelf is de optie "CHANGE PASSWORD" in het hoofdmenu). Mogelijke instellingen zijn "Disabled", "Setup" (alleen de toegang tot de BIOS SETUP is beveiligd) en "Always" (bij het starten van de PC).
- **Video, adaptor en system shadow**
  Deze instellingen kunnen ingedeeld worden in drie hoofdgroepen:
  - "Video ROM Shadow C?00,16K";
  - "Adaptor ROM Shadow ??00,16K" ;
  - "System ROM Shadow F000,64K".

Het "schaduwen" is bedoeld om de verwerkingssnelheid van de PC te verhogen door de inhoud van trage EPROM-geheugens te dupliceren op bepaalde gereserveerde adressen in het snellere interne RAM-geheugen.

19.10 De opstart-procedure van de PC



Figuur 3/19.10-19: De geavanceerde chipset SETUP.

De inhoud is dan in het RAM-geheugen veel sneller beschikbaar. Belangrijk zijn vooral het "schaduwen" van het ROM BIOS van de grafische kaart (Video ROM Shadow C?00,16K) en van de PC zelf (System ROM Shadow F000,64K). Zet deze drie posities bij voorkeur op "Enabled". Het kan echter gebeuren dat bepaalde geheugensegmenten door andere adapter-kaarten in gebruik zijn. Een voorbeeld is bijvoorbeeld een zogenoemde "LTP-2 printer". Een dergelijke printer heeft een eigen kaart, die een deel van het geheugen boven de 640 kB grens in gebruik heeft. Het kan zijn dat dit geheugengebied ook wordt gebruikt bij het shadowen en dan ontstaan geheugenconflicten, waardoor de printer niet meer zal werken. Alvorens men deze opties dus op "Enabled" zet is het verstandig door middel van een programma dat het geheugen kan onderzoeken, bijvoorbeeld "MSD" van MS-DOS 6.x, te onderzoeken welke segmenten boven de 640 kB door kaarten worden bezet.

De Geavanceerde chipset SETUP

Het beeldscherm van de geavanceerde chipset SETUP (ADVANCED CHIPSET SETUP) die vanuit het hoofdmenu toegankelijk is, zal in veel opzichten lijken op de weergave in figuur 3/19.10-19. Afhankelijk van de versie zijn echter ook hier weer afwijkingen mogelijk. Deze chipset-opties betreffen voornamelijk instellingen van de elektronica op het moederbord van de PC. Van wezenlijk belang voor de prestaties van het systeem is vooral de snelheid waarmee in het intern geheugen kan worden gelezen en geschreven en de klokfrequentie van het bus-systeem. Ook nu volgt een korte beschrijving van de functie van de diverse opties:

- **BUS clock selection**
  Een belangrijke instelling is die van de klokfrequentie van de bus op het moederbord die onder meer de snelheid van de gegevensoverdracht tussen uit-

breidingskaarten en het overige systeem bepaalt. Mogelijke waarden worden aangegeven in de vorm "CLKIN/3"; deze waarde zet bijvoorbeeld de bus klokfrequentie op een derde van de klokfrequentie van de centrale processor. Afhankelijk van de BIOS-versie zijn waarden mogelijk tussen "CLKIN/2" en "CLKIN/8".

Fabrikanten van uitbreidingskaarten bevelen de volgende instellingen voor een optimale werking van het systeem aan:

- **DRAM write/read wait state**
  Bij de opties "DRAM Write Wait State" en "DRAM Read Wait State" bestaat de mogelijkheid het aantal wachtcycli (Wait State) voor het schrijven respectievelijk het lezen in het intern geheugen (DRAM, Dynamic RAM) vast te leggen. Mogelijke waarden liggen tussen "0 W/S" en "4 W/S". Hierbij geldt natuurlijk in eerste instantie het devies: hoe minder wachtcycli, hoe sneller. De juiste keuze van het aantal wachtcycli is echter afhankelijk van het type van de aanwezige geheugen-chip's en de klokfrequentie van de centrale processor in het systeem. Helaas is de snelste instelling van "0 W/S" met name bij PC's met een hoge klokfrequentie niet altijd haalbaar. Als men een "W/S"-waarde kiest die het systeem niet aankan, zal dit meestal al bij het starten van de PC blijken doordat de PC simpelweg gaat "hangen". Men moet dan via de SETUP meer wachtcycli instellen.
- **Cache write/read wait state**
  In sommige versies is het tevens mogelijk de snelheid van het lezen en schrijven in het cache-geheugen door middel van wachtcycli te regelen. Men doet er echter verstandig aan de voorgestelde waarden ongewijzigd over te nemen!
- **Non-cacheable block-? base/size**
  Met deze opties kunnen bepaalde geheugengebieden worden uitgesloten van het gebruik als cache-geheugen. Over het algemeen komt het de snelheid van het systeem ten goede als het over zoveel mogelijk cache-geheugen kan beschikken. Wijzig de instellingen derhalve alleen als dat bijvoorbeeld ten gevolge van geheugenconflicten van uitbreidingskaarten echt noodzakelijk is.

Overige opties van de SETUP

Het hoofdmenu geeft voorts toegang tot andere opties van het AMI BIOS SETUP programma zoals onder andere een automatische configuratie met BIOS standaardinstellingen (AUTO CONFIGURATION WITH BIOS DEFAULTS) of een automatische configuratie met de instellingen zoals ze bij het inschakelen van de PC golden (AUTO CONFIGURATION WITH POWER-ON DEFAULTS).

Bij de automatische configuratie met behulp van standaard BIOS-instellingen zorgt het SETUP-programma zélf voor bruikbare, voor een deel zelfs individuele instellingen van het systeem.

Als men zich om gedetailleerde instellingen, zoals hiervoor beschreven, liever niet wil bekommeren, dan kan men gerust van deze automatische configuratie gebruik maken.

Minder gunstig is het gebruik van de automatische configuratie op basis van de waarden bij het inschakelen van de PC. Na deze configuratie zal over het algemeen de PC weliswaar moeiteloos werken, de instellingen zijn echter niet optimaal afgestemd op de mogelijkheden van het systeem.

Dientengevolge zullen de prestaties achterblijven bij de optimale instellingen.

19.10 De opstart-procedure van de PC

Verdere opties in het hoofdmenu betreffen onder andere het wijzigen van het wachtwoord (CHANGE PASSWORD). Om de wachtwoord beveiliging in werking te laten treden, moet wel de betreffende instelling in de standaard CMOS SETUP ("Password Checking Option") geactiveerd zijn. Vóór het wijzigen van het wachtwoord wordt eerst gevraagd het tegenwoordig geldende wachtwoord op te geven. Dat weet men natuurlijk niet als men voor het eerst van deze mogelijkheid gebruik maakt. Mocht de documentatie bij de PC geen andersluidend wachtwoord noemen, dan is het wachtwoord "AMI" een goede gok voor het initiële wachtwoord.

Hard disk utility

Via de optie "HARD DISK UTILITY" kan men de vaste schijf of schijven inrichten (onder meer low level formatteren). Hier is echter voorzichtigheid geboden: ***als het om een vaste schijf van het type AT-bus of IDE gaat, moet het gebruik van dit hulpprogramma sterk worden ontraden.*** Dergelijke vaste schijven zijn in de fabriek reeds low level geformatteerd. Als men dit opnieuw doet, kunnen de schijven zelfs beschadigd raken zonder dat men aanspraak kan maken op een eventueel nog geldende garantie.

3/19.11

Principes en werking van printers

Inleiding

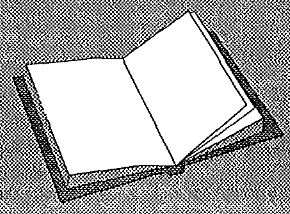
Technologieën
In de loop der jaren zijn er verschillende technologieën ontwikkeld, die één eigenschap gemeen hebben. Zij zorgen ervoor dat digitale gegevens, onder de vorm van bits opgeslagen in een computer, worden omgezet in leesbare uitvoer op papier. In dit hoofdstuk zullen al deze technologieën worden besproken, met hun typische eigenschappen, hun voor- en nadelen. Hierbij worden zowel monochrome als kleurenprinters behandeld.

Dertien technologieën
Op dit moment zijn dertien verschillende technologieën beschikbaar om aan het vooropgestelde doel te voldoen:
- monochrome naaldmatrix printers;
- kleuren naaldmatrix printers;
- monochrome thermo-direct printers;
- monochrome thermo-transfer printers;
- kleuren thermo-transfer printers;
- monochrome inkjet printers;
- kleuren inktjet printers;
- monochrome laserprinters;
- kleuren laserprinters;
- kleuren thermal wax printers;
- kleurstof sublimatie printers;
- ionen printers;
- nat-elektrostatische kleurenprinters.

Gemeenschappelijke eigenschap
Alle printers hebben echter één gemeenschappelijke eigenschap: de matrixtechniek. Dit is een rechtstreeks gevolg van het feit dat computers binair werken. Computers kennen slechts twee data-toestanden: "L" of "H". Ook de gegevens die geprint moeten worden, bestaan dus alleen maar uit "L"-en en "H"-en. Het gevolg is dat iedere printer slechts twee printtoestanden kent: ofwel wordt er iets op het papier geprint, ofwel wordt er niets op het papier geprint. Iedere printer drukt dus in wezen alleen maar puntjes af. Weliswaar zijn bepaalde technologieën in staat deze welles/nietes-gegevens ietsjes uit te smeren of af te vlakken, maar dat verandert niets aan deze fundamentele eigenschap van printers.

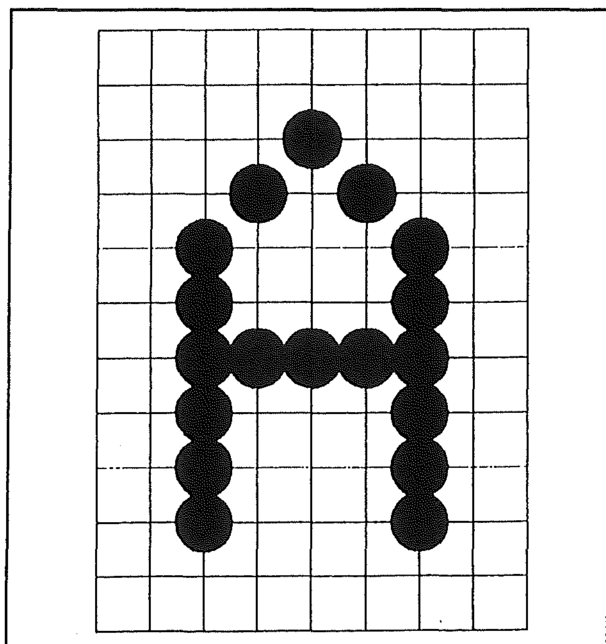
Het matrix-principe
Wil men bijvoorbeeld een letter printen, dan zal men de in wezen vloeiende vorm

LEES OOK:
Geen verwijzingen



19.11 Principes en werking van printers

van de letter moeten omzetten in iets dat bestaat uit een aantal puntjes.
Dat kan door gebruik te maken van het matrix-principe. Dit principe wordt toegelicht aan de hand van figuur 3/19.11-1. De vloeiende vorm van de letter wordt geprojecteerd op een lijnenmatrix. Ieder kruispunt van een vertikale met een horizontale lijn correspondeert met een punt, dat al dan niet geprint kan worden. Een dergelijk punt noemt men een "dot", het allerkleinste inktspatje dat de printer op papier kan zetten. De dotjes, die de vorm van de letter bedekken, moeten geprint worden, alle overige punten blijven blank.



Figuur 3/19.11-1: Het algemene principe van printers: afdrukken van dots op de kruispunten van een lijnenmatrix.

In het voorbeeld wordt de letter A gevormd uit een matrix van tien horizontale lijnen bij zeven vertikale lijnen. In principe heeft deze printer dus zeventig dots ter beschikking voor het vormgeven van alle te drukken letters en tekens.

De resolutie
Een rechtstreeks gevolg van deze algemene eigenschap is dat alle printers gekarakteriseerd worden door een belangrijke eigenschap: hun resolutie. Deze eigenschap definieert de fijnheid, waarmee de gegevens kunnen worden afgedrukt. Of, met andere woorden, het beschikbaar aantal vertikale en horizontale lijnen van de matrix per lengte-eenheid. De resolutie wordt uitgedrukt met de eenheid "dpi", afkorting van de Engelse uitdrukking "dots per inch". Deze eenheid definieert het aantal horizontale en vertikale lijnen van de matrix. Een printer die een resolutie heeft van 300 dpi kan dus zowel horizontaal als vertikaal 300 dots per inch (2,54 cm) op papier zetten. Let wel dat het niet per definitie zo is, dat de horizontale en vertikale resoluties aan elkaar gelijk zijn. Er zijn printers, die bijvoorbeeld een horizontale resolutie van 300 dpi hebben en een vertikale van 600 dpi.
Wat de resolutie van printers betreft, lijkt het er op alsof er geen technologische grenzen meer bestaan. Waren jarenlang de 360 dpi van matrixnaald printers en de 300 dpi van laserprinters het hoogst bereikbare, sinds enige jaren gaan die cijfers met sprongen omhoog. Laserprinters en inkjet printers met een resolutie van 600 dpi zijn op dit moment standaard. Kleurenprinters zijn nu zelfs betaalbaar leverbaar met een resolutie van 720 dpi. LaserMaster heeft onlangs een A3 laserprinter op de markt gebracht met de ongelooflijk hoge resolutie van 1.800 dpi! Hierdoor vervagen de grenzen tussen kantoorprinters en de zeer dure fotografische belichters.

Matrix versus FFC
In de fundamentele matrix-eigenschap onderscheiden alle printers zich van het

19.11 Principes en werking van printers

principe waarmee oude schrijfmachines werken. Schrijfmachines voldoen aan het FFC-principe, afkorting van "Fully Formed Character". Bij deze apparaten wordt gebruik gemaakt van hamertjes met stansvormen, die via een inktlint tegen het papier geslagen worden. Deze stansvormen hebben vloeiende omrandingen, die precies het profiel van een letter volgen. Oftewel, de stansvormen zijn volledig gevormd naar het te printer karakter. Dergelijke systemen hebben dus een in principe oneindig hoge resolutie.
Hoewel het FFC-principe als voordeel heeft dat de gedrukte tekst heel mooi is, is dit ene voordeel door de moderne printertechnologieën volledig achterhaald. Matrixen met een resolutie van 1.000 dpi zijn tegenwoordig betaalbaar beschikbaar, zodat men een microscoop nodig heeft om vast te stellen dat de afgedrukte letters niet vloeiend van vorm zijn, maar samengesteld zijn uit dots. Bovendien heeft het matrixprincipe als groot voordeel dat men er alles kan mee afdrukken. Met het FFC-principe kan men alleen die karakters en symbolen afdrukken die beschikbaar zijn op de hamertjes. Met matrixprinters kan men niet alleen karakters printen, maar ook tekeningen, grafieken en zelfs foto's. In principe is immers alles dat zichtbaar is samen te stellen met een puntenmatrix.

Klassen

Printers kunnen, naar gelang hun werkingsprincipe, ingedeeld worden in een aantal klassen:
- regel of pagina;
- impact of non-impact.

Deze eerste klasse houdt verband met de manier waarop gegevens op het papier worden gezet. Er zijn printers die dat regel voor regel doen en er zijn printers die in een keer een volledige pagina bedrukken. Bijgevolg kan men printers indelen in regelprinters en paginaprinters.
Bij *regelprinters* wordt de te drukken informatie regel na regel van de computer naar de printer verstuurd. Na ontvangst van de regelinformatie zal de printer deze ene regel afdrukken. Nadien wordt het papier een regelhoogte getransporteerd en is de printer klaar voor het verwerken van de volgende regel. Eigenschap van regelprinters is dus dat het papier stapsgewijs door de printer wordt getransporteerd.
Bij *paginaprinters* worden eerst alle gegevens van één pagina naar de printer verstuurd. De elektronica in de printer zet die informatie om naar matrixgegevens, die in het geheugen worden opgeslagen. Eerst nadat de volledige pagina is omgezet in een matrixbeeld, wordt dit afgedrukt op het papier. Het papier wordt in één continue beweging door de printer gevoerd en de informatie wordt in eenzelfde vloeiende beweging overgedragen op het papier.
Een tweede indelingsklasse houdt rekening met de manier waarop de printer de gegevens op het papier zet. Er zijn printers, waarbij het mechanisme van het apparaat door middel van mechanische kracht contact met het papier maakt. Dit noemt men *impact* printers. Anderzijds zijn er printers, waarbij geen of nauwelijks mechanische kracht wordt gebruikt bij het contact tussen het mechanisme en het papier. Uiteraard noemt men deze apparaten *non-impact* printers. Soms is er zelfs helemaal geen contact tussen het printmechanisme en het papier.
Het grote voordeel van impact printers is dat zij in staat zijn doorslagen te printen. Dat kan door middel van het ouderwetse carbonpapier, maar ook door gebruik te maken van de moderne doordruk papier-

soorten. Vooral in het bedrijfsleven is dit nog steeds een belangrijk voordeel! Daar immers heeft men meestal één of meerdere copieën nodig van alles dat geprint wordt. Het is veel handiger en goedkoper deze in één keer door een printer te laten bedrukken, dan voortdurend naar het copieermachine te lopen!

De printsnelheid

Een andere belangrijke eigenschap is de snelheid waarmee geprint wordt. Men heeft pogingen ondernomen om hiervoor een internationaal geaccepteerde referentiestandaard op te stellen.
Voor regelprinters gaat dat eenvoudig. Hierbij wordt gewoon het maximaal aantal karakters opgegeven, dat per seconde geprint kan worden. Als eenheid wordt de cps gekozen, afkorting voor het Engelse "characters per second". Uiteraard is deze snelheid van een aantal instellingsfactoren afhankelijk, zoals of men kiest voor lage of hoge kwaliteit afdrukken. Bij de specificatie van het aantal cps moet dit dus als parameter opgegeven worden.
Voor paginaprinters ligt het definiëren van de snelheid moeilijker. In principe wordt het aantal pagina's opgegeven, dat de printer per minuut kan bedrukken. Als eenheid wordt de ppp gekozen, afkorting van "pages per minute". Maar hoeveel pagina's een printer per minuut kan bedrukken hangt volledig af van wat er op die pagina staat. Vandaar dat de Europese computerfabrikanten, verenigd in de "European Computer Manufacturers Association" (ECMA), een standaard hebben afgesproken voor het ondubbelzinnig definiëren van de snelheid van paginaprinters. Deze standaard gaat door het leven onder de naam ECMA-132. Deze standaard werd in 1990 overgenomen door de "EPMI", de "European Printer Manufacturers and Importers", zodat het op dit moment zo zou moeten zijn dat alle in Europa gemaakt en verkochte printers met elkaar vergeleken kunnen worden.
Bij de ECMA-132 test moet iedere printer drie documenten vijf maal drukken:
- een standaard zakenbrief;
- een blad uit een spreadsheet;
- een eenvoudige grafiek.

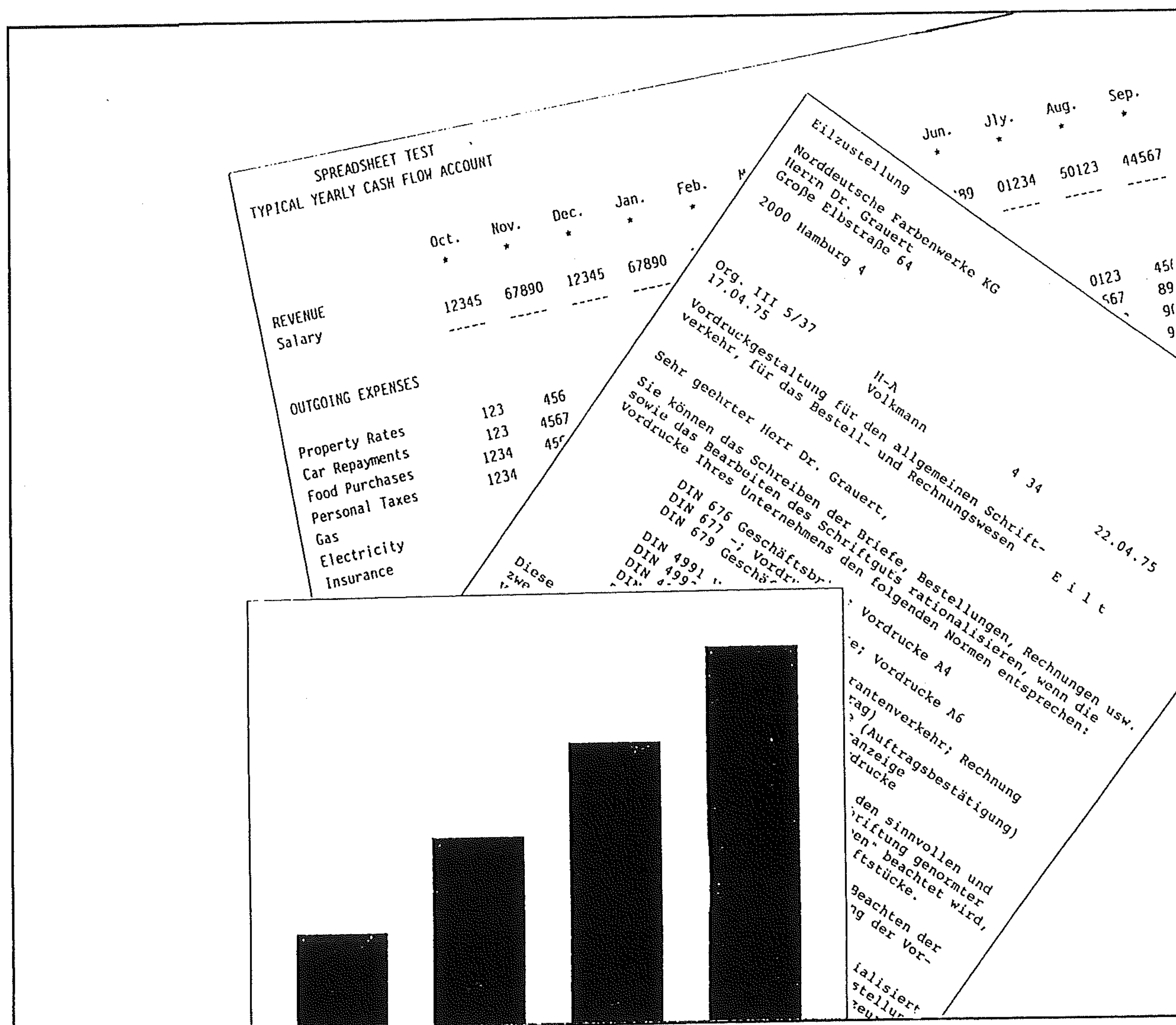
Deze drie standaard brieven zijn voorgesteld in figuur 3/19.11-2. Nadien wordt een gemiddelde snelheid berekend en uitgedrukt in pagina's per minuut. Wie deze figuur bekijkt, zal vaststellen dat de drie ECMA- pagina's nu niet bepaald voldoen aan de moderne kwaliteitseisen die aan printwerk gesteld worden! Bovendien bepalen op dit moment allerlei ingewikkelde softwarematige printerdrivers voor een belangrijk deel de snelheid van het apparaat. Vandaar dat deze verouderde testnorm op dit moment zeer omstreden is.

De printertaal

Een printer wordt op een computer of een netwerk aangesloten via een gestandaardiseerd hardware-protocol, zoals:
- serieel RS-232;
- parallel Centronics;
- LocalTalk;
- Ethernet.

Die protocollen zijn goed beschreven en zullen nauwelijks voor problemen zorgen. Daarnaast moet iedere printer echter ook softwarematig met de PC kunnen communiceren. Hiervoor is een softwarematig protocol nodig en de ellende is dat er hiervan in de loop der jaren heel wat ontwikkeld zijn! Bovendien verschijnt er versie na versie van een en hetzelfde protocol, die ieder op zich weer veel meer mogelijkheden hebben dan de vorige versie.

19.11 Principes en werking van printers



Figuur 3/19.11-2: De drie testpagina's die volgens de ECMA-132 standaard vijf maal door paginaprinters afgedrukt moeten worden.

De meeste protocollen zijn zogenoemde "pagina beschrijvingstalen". Dat betekent dat het protocol een aantal instructies bevat, waarmee alles dat op het te drukken vel papier staat heel precies beschreven wordt. Zo'n protocol gaat door het leven onder de naam "printertaal" en de meest gebruikte standaarden of de-facto standaarden zijn:

- **ESC/P**, een reeds in 1981 door Epson ontwikkelde taal, die gebruikt wordt voor het aansturen van regelprinters en géén pagina beschrijvingstaal is, maar alleen besturingscommando's bevat.
- **HPGL**, een door Hewlett-Packard ontwikkelde besturingstaal voor wetenschappelijke plotters, waarvan tegenwoordig echter ook een versie bestaat die paginaprinters kan aansturen.
- **PCL**, een de-facto standaard die door hetzelfde bedrijf werd geïntroduceerd bij hun eerste laserprinters en inmiddels door zowat alle fabrikanten van laserprinters wordt ondersteund.

- **Prescribe**, een eigenzinnige taal die door Kyoceras werd ontwikkeld en een beetje op BASIC lijkt.
- **Postscript**, een door Adobe ontwikkelde pagina beschrijvingstaal, die eerst in de drukkerswereld een standaard werd en nu ook voor een heleboel goedkopere printers leverbaar is. Vooral versie 2.0 van Postscript is een de-facto standaard voor kleurenprinters die volgens het pagina-principe werken.
- **TrueImage**, een door MicroSoft ontwikkeld antwoord op Postscript, waarmee gepoogd werd de overheersing van Adobe te doorbreken en dat voornamelijk bekend staat door de TrueType-letters die er deel van uitmaken.
- **GDI**, de grafische interface die standaard deel uitmaakt van Windows 3.1 en Windows 95 en waarmee printers rechtstreeks vanuit de computer aangestuurd kunnen worden.

Op het einde van dit hoofdstuk worden deze printertalen uitgebreider behandeld.

Rastersystemen en kleurenschema's
Tot voor enkele jaren werden zelfs de beste printers alleen ingezet voor het drukken van tekst en eenvoudige lijntekeningen en grafieken. Iedere poging om een foto, elektronisch bewaard onder de vorm van een TIF-, GIF-, BMP- of PCX-bestand, op een printer af te drukken, leverde een resultaat op dat zelfs met de beste wil van de wereld niet beter te omschrijven was als "volstrekt waardeloos". Nu echter de resolutie van printers stormachtig stijgt en snelle processoren toelaten zeer uitgebreide software-algorithmen toe te passen, komt het afdrukken van foto's binnen ieders bereik.
Een foto is een halftoon beeld, niet opgebouwd uit dots maar uit een continu helderheids- en/of kleurverloop. In een zwart/wit-foto kan men honderden grijstinten aantreffen. Een printer kan echter alleen zwart afdrukken en dus moeten er methodes verzonnen worden om grijstinten na te maken door middel van zwarte puntjes, geplaatst in een bepaald patroon. Iedere grijstint komt overeen met een specifiek patroon van dotjes, geplaatst in een kleine matrix. Doordat ons oog de zeer kleine dotjes niet individueel kan waarnemen, ontstaat een optische indruk van een grijs vlak. Dit algemeen procédé noemt men "rasteren" en dat wordt al zeer lang toegepast bij het afdrukken van foto's in kranten en tijdschriften. De laatste vijf jaar staan ook op dit gebied de ontwikkelingen niet stil. Er zijn talrijke nieuwe rasterprocédés ontwikkeld, speciaal bedacht om op de beste manier halftoon afbeeldingen af te drukken op bijvoorbeeld laserprinters. Namen en begrippen, die op dit gebied standaarden hebben gezet, zijn:
- Bayer dithering;
- Floyd-Steinberg dithering;
- Stucki dithering;
- amplitudegemoduleerd rasteren;
- frequentiegemoduleerd rasteren;
- stochastisch rasteren.

Ook aan deze zeer interessante technieken wordt aan het einde van dit hoofdstuk de nodige aandacht besteed.

Monochrome naaldmatrix printers

Inleiding
Hoewel naaldmatrix printers verouderd zijn, wordt hier toch veel aandacht besteed aan hun constructie, werking en

19.11 Principes en werking van printers

eigenschappen. De naaldmatrix printers zijn immers hét model, waarnaar moderne printers als thermo-transfer printers en zelfs inkjet printers zijn vormgegeven.

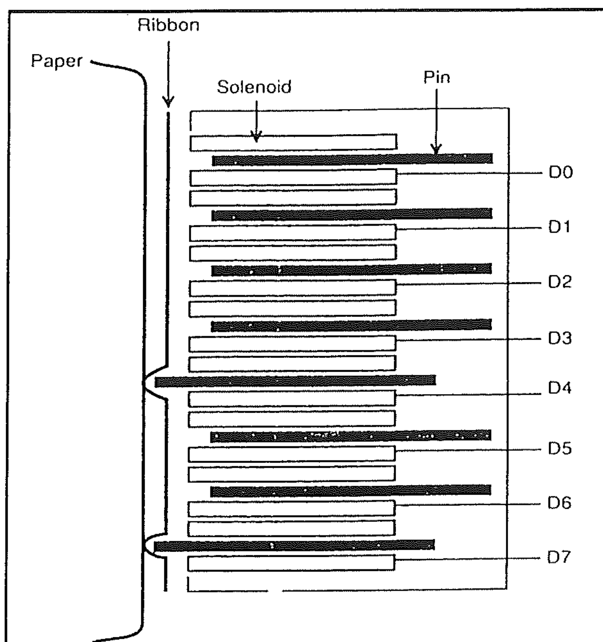
Het principe
Hoewel in principe alle printers volgens het matrixprincipe werken, verstaat men onder het begrip "naaldmatrix printer" een printer waarbij het drukbeeld tot stand komt door een aantal naaldjes via een inktlint op het papier te laten slaan. Het papier wordt getransporteerd middels een rubber rol, die er voor zorgt dat de naaldjes goed contact maken met het papier. Het zal dus duidelijk zijn dat matrix printers impact printers zijn. De naaldjes zitten in een printkop en zij worden elektromagnetisch aangedreven. Matrix printers met naalden heten officieel SIDM-printers, afkorting van "Serial Impact Dot Matrix".

Regelprinter
Een naaldmatrix printer is een regeldrukker. De naaldjes zitten in een printkop, die door middel van een geleidingsmechanisme van links naar rechts of van rechts naar links over het papier wordt verplaatst. Tezelfdertijd wordt de informatie van deze regel op het papier geprint. Nadien zorgt de rubber transportrol ervoor dat het papier over één regel wordt verplaatst. Ook bij het afdrukken van grafische afbeeldingen wordt op deze manier gewerkt. De software in de printerdriver ontleedt het volledige beeld in regel-na-regel informatie en stuurt deze informatie regelgewijs naar de printer.

Aantal naalden
Een belangrijke specificatie van naaldmatrix printers is het aantal naalden waarover zij beschikken. De allereerste modellen hadden 8 of 9 naaldjes, de moderne uitvoeringen bezitten allemaal 18 tot 24 naaldjes. Het zal duidelijk zijn dat het aantal naaldjes de maximale resolutie van de printer bepaalt. Toch is het niet zo dat men zonder meer kan stellen dat de kwaliteit van het drukbeeld rechtstreeks afhankelijk is van het aantal naaldjes. Het probleem is dat de naaldjes natuurlijk een bepaalde onderlinge afstand hebben. Anderzijds hebben zij een bepaalde dikte, die varieert van 0,1 tot 0,4 mm. Er bestaat een bepaalde verhouding tussen naaldafstand en naalddikte, waarbij een optimaal letterbeeld ontstaat. Deze optimale verhouding kan technologisch uitstekend gerealiseerd worden bij printkoppen met 18 naaldjes. Bij de 24-naald modellen is dat moeilijker. In de praktijk zijn daar de naaldjes iets te dun, waardoor een rafelig letterbeeld ontstaat. Voor het Europese schrift zijn 18 naaldjes trouwens meer dan voldoende. De printkoppen met 24 naaldjes zijn een Japanse ontwikkeling, omdat bleek dat 18 naalden niet genoeg waren om de ingewikkelde letters van het Japanse schrift keurig te vormen.

De constructie van de printkop
Zoals uit figuur 3/19.11-3 blijkt, zitten de naalden op een rijtje (in-line) in de printkop. Dit geldt voor 9-naald printers, bij 18- en 24-naalden wordt gebruik gemaakt van twee rijen van respectievelijk 9 en 12 naalden, die onderling iets verschoven zijn (dual in-line staggered). De onderlinge afstand van de printnaalden moet voldoen aan een ECMA-standaard. De witruimte tussen de geprinte dots mag niet groter zijn dan 0,02 mm. Een en ander is het gevolg van het feit dat het schrift gelezen moet kunnen worden door OCR-apparatuur, bijvoorbeeld bij de volledig geautomatiseerde verwerking van giro-

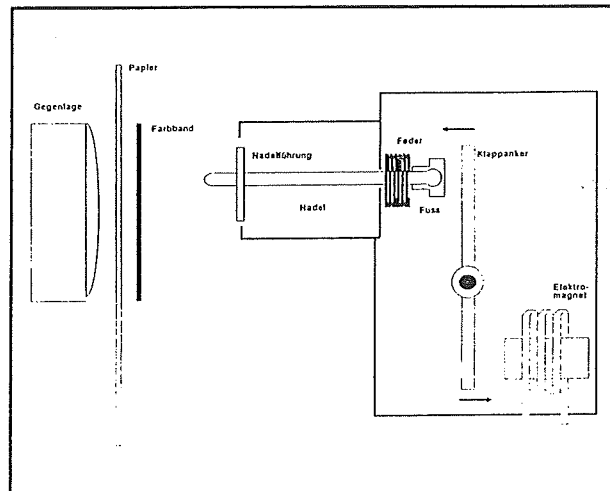
kaarten. Iedere naald wordt aangestuurd door een uiterst klein elektromagneetje. Als dit spoeltje onder spanning wordt gezet zal het magnetisch veld er voor zorgen dat het naaldje met kracht uit de printkop schiet. Het naaldje drukt dan het inktlint tegen het papier, waardoor er één dotje wordt gedrukt.



Figuur 3/19.11-3: Principiële werking van de printkop van naaldmatrix printers.

De ballistische balansanker printkop

In figuur 3/19.11-4 is de meest gebruikte constructie van de printkop getekend. Deze kop werkt volgens het "ballistische balansanker"-principe. De naald kan vrij bewegen in een ceramische houder en wordt door middel van een klein veertje in de rustpositie gedwongen. Het interne uiteinde van de naald is uitgerust met een klein schokdempertje van rubber. De naald wordt geactiveerd door een balansanker, dat draait rond een asje en aan de onderzijde wordt aangetrokken door het magneetveld van de spoel van de elektromagneet. De elektromagneet kan een flinke stroom verdragen.



Figuur 3/19.11-4: De constructie van de ballistische balansarm printerkop.

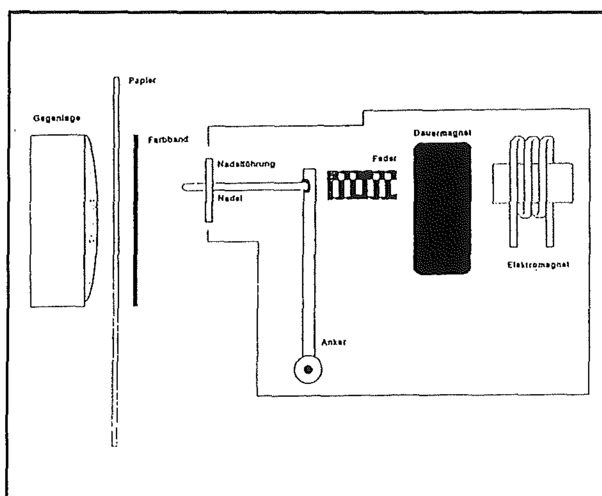
De kracht, waarmee de naald tegen het inktlint slaat, is in te stellen door de grootte van het magneetveld te wijzigen. Op deze manier kan men met dergelijke printkoppen tien goed leesbare doorslagen maken. In de vakwereld noemt men dit HFP, letterwoord voor "Heavy Forms Printing". Een nadeel van deze constructie is dat er heel wat elektrisch vermogen in de spoeltjes gepompt moet worden. Er moet immers een sterk magneetveld ontstaan om de naaldjes met grote kracht weg te schieten. Een en ander heeft tot gevolg dat een ballistische kop behoorlijk heet kan worden. De maximaal toelaatbare hitte-ontwikkeling stelt bepaalde grenzen aan het maximaal aantal naaldbewegingen per seconde, met andere woorden aan de maximale printsnelheid.

De stored-energy printkop

De constructie van een dergelijke printkop is getekend in figuur 3/19.11-5. De naald is gekoppeld aan een draaibaar opgesteld anker. Achter het anker zit een

19.11 Principes en werking van printers

klein veertje, dat door een sterke permanente magneet in de samengedrukte rustpositie wordt gehouden. Achter de permanente magneet staat een spoeltje. Als dit spoeltje onder spanning wordt gezet zal het elektromagnetisch veld van het spoeltje het permanent magnetisch veld van de magneet verzwakken. Het resulterende veld is niet in staat de veer in de samengedrukte positie te houden. De veer ontspant zich, slaat tegen het anker en het anker drukt de naald tegen het inktlint.



Figuur 3/19.11-5: De constructie van een printkop die werkt volgens het "stored-energy"-principe.

De naam van deze koppen is te verklaren uit het feit dat de energie, noodzakelijk om de naald met voldoende kracht tegen het inktlint te drukken, nu niet alleen geleverd moet worden door de elektrische energie die men naar het spoeltje stuurt. Een groot deel van de energie zit in de krachtenbalans tussen de samengedrukte veer en de permanente magneet. Het volstaat deze krachtenbalans met een klein stroompulsje te verbreken om het systeem in werking te zetten.

Het voordeel van de "stored-energy"-constructie is dat er veel minder elektrische energie in de printkop geïnvesteerd moet worden. Deze wordt dus minder heet, met als gevolg dat de printsnelheid aanmerkelijk verhoogd kan worden. Een moderne "stored-enegy"-kop kan een snelheid van 1.800 Hz bereiken, hetgeen betekent dat een naaldje maximaal 1.800 keer per seconde geactiveerd kan worden!

De volledige constructie

In figuur 3/19.11-6 is een exploded view getekend van een moderne naaldmatrix printer. Men herkent de dikke metalen staaf, waarlangs de printkop heen en weer kan glijden.

De printkop is door middel van een flexibele vlakke kabel verbonden met de elektronica. Zowel de printkop als de rubber rol waar het papier omheen draait worden aangedreven door zeer nauwkeurig werkende stappenmotoren. Bedenk dat beide mechanismen bewegingen van tienden van een millimeter moeten kunnen maken!

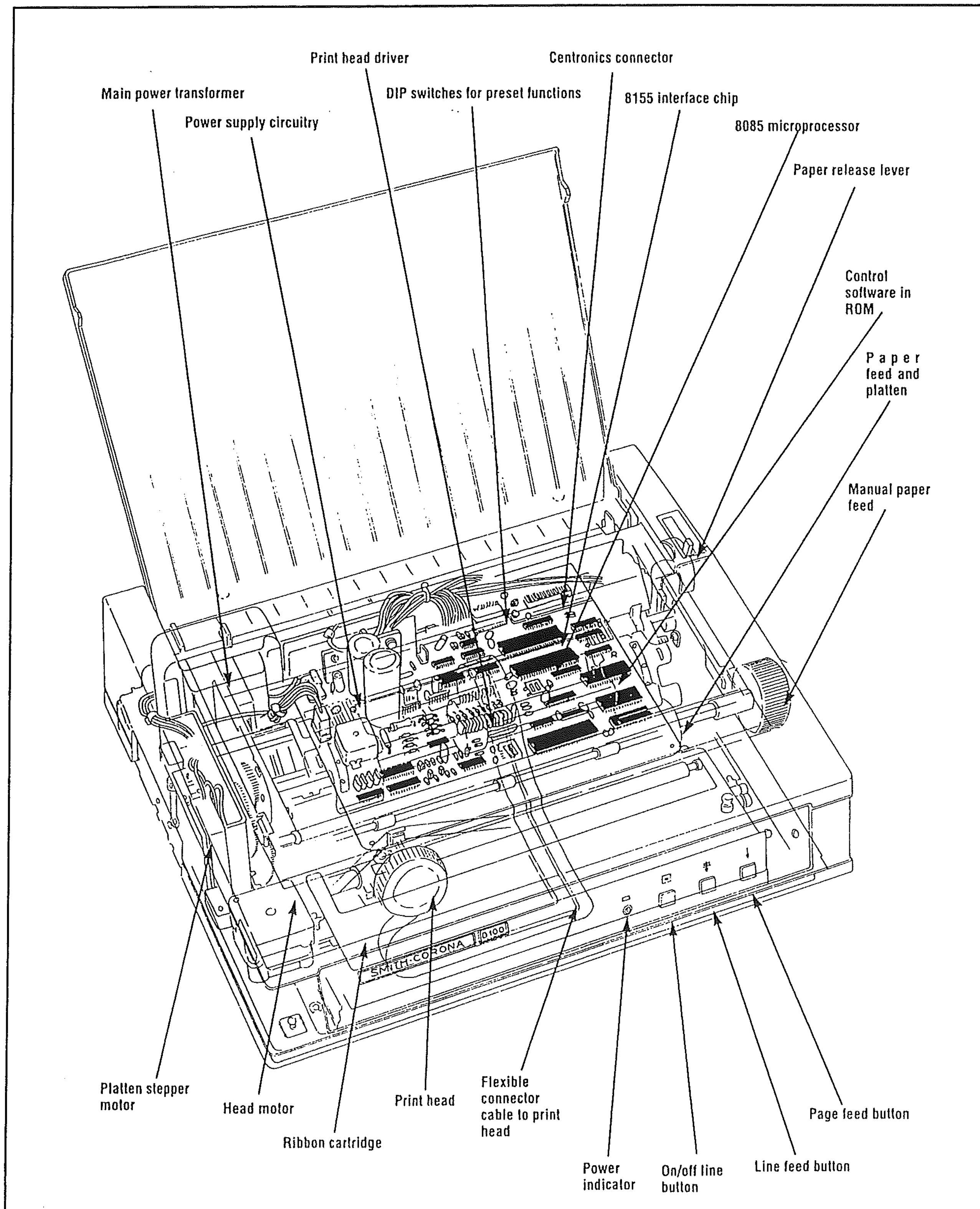
De resolutie van naaldmatrix printers

De maximale resolutie van naaldmatrix printers is afhankelijk van het aantal naalden en de manier waarop deze naalden worden aangestuurd:

- met 9-naald printers kan men een resolutie bereiken tussen 60 en 240 dpi;
- met 24-naald modellen een resolutie tussen 240 en 360 dpi.

Bij deze resolutie moet men echter wél bedenken dat 24-naald printers in de letter-modus slechts gebruik maken van 18 naalden! De 24 naalden worden alleen allemaal ingeschakeld als men de printer in de grafische modus zet en grafieken of tekeningen verwerkt.

19.11 Principes en werking van printers

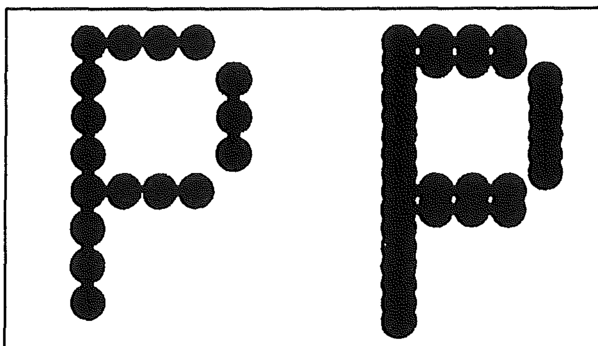


Figuur 3/19.11-6: De constructie van een moderne naaldmatrix printer.

19.11 Principes en werking van printers

DPQ en NLQ

Men zou denken dat een bepaald aantal naalden overeen komt met een bepaalde resolutie. Dat is ook zo, maar men heeft een middel verzonnen om de resolutie van een naaldmatrix printer schijnbaar te verbeteren. Hierbij komen de twee begrippen "DPQ" en "NLQ" om de hoek kijken. Het verschil wordt toegelicht aan de hand van figuur 3/19.11-7. Bij snelle DPQ-druk (afkorting van "data processing quality") wordt de printkop bij het printen van een regel één keer van links naar rechts of van rechts naar links over het papier gevoerd. De naaldjes drukken hun matrix af en het gevolg is een drukbeeld met lage resolutie (links in de tekening).



Figuur 3/19.11-7: Het verschil tussen DPQ-druk (links) en NLQ-druk (rechts).

Bij NLQ-druk (afkorting van "nearly letter quality") wordt de printkop twee keer over de regel gevoerd, waarbij het papier vóór de tweede doorgang iets verschoven wordt. Het gevolg is dat de dotjes van de eerste doorgang worden aangevuld met een tweede serie dotjes, die er precies tussenin liggen (rechts in de tekening). Het rafelige letterbeeld van de randen van de letters wordt daardoor recht getrokken, zodat een beter resultaat ontstaat met een schijnbaar hogere resolutie. De letters worden echter in één richting onnatuurlijk dik. Het nadeel van NLQ-druk is uiteraard dat de printsnelheid gemiddeld genomen de helft lager wordt.

Afdruksnelheid

Naaldmatrix printers kunnen een maximale snelheid halen van 400 cps in de DPQ-mode. Dit wordt niet alleen bepaald door het soort printkop dat wordt toegepast, maar ook door de intelligentie van het besturingssysteem. Er zijn printers die niet te bedrukken delen van een regel met hoge snelheid overslaan en die niet steeds na het einde van een regel terug gaan naar de linker kantlijn, maar de volgende regel van rechts naar links afdrukken.

Voordelen

Het grote voordeel van de naaldmatrix printer is dat het een impact printer is, zodat men doorslagen kan maken. Het is zelfs zo, dat de naaldmatrix printer het enige apparaat is waarmee dit kan! Bovendien heeft de printer geen moeite met kettingpapier, zodat dergelijke printers nog steeds in kantooromgevingen worden toegepast om met grote snelheid duizenden rekeningen per uur af te handelen.

Nadelen

Het grote nadeel van een impact printer is dat het apparaat zeer veel lawaai produceert. De uiterst dunne naaldjes worden met een grote kracht tegen het inktlint en het papier geslagen hetgeen, gecombineerd met de hoge frequentie van de naaldbewegingen, een zeer storend ratelend geluid tot gevolg heeft.
Een tweede nadeel is dat dergelijke apparaten nogal storingsgevoelig zijn en onderhoud vergen. Het inktlint strijkt voortdurend tegen de printkop, waardoor deze laatste langzaam maar zeker vervuilt. Ge-

droogde inktresten kunnen een gaatje, waardoor een naaldje naar buiten wordt geslagen, gedeeltelijk afsluiten. Hierdoor kan een naald vastlopen of zelfs breken. Het gevolg is dat er door iedere geprinte regel een duidelijk herkenbare witte streep loopt. Naalden zijn niet te vervangen, het enige dat er op zit is een volledig nieuwe printkop te monteren.

Kleuren naaldmatrix printers

Gekleurd lint

Naaldmatrix printers zijn in principe ook geschikt voor het afdrukken van kleuren. Het zwarte inktlint wordt dan vervangen door een inktlint dat bestaat uit vier kleurbanden met als kleuren zwart, rood, blauw en geel. Het lintgeleidingsmechanisme is nu bestuurbaar, zodat de software een van de kleurbanden voor de naaldjes kan zetten. Met een dergelijk lint kunnen, door een eenvoudige additieve menging van de beschikbare basiskleuren, in principe zeven kleuren worden afgedrukt. In principe, omdat niet alle mengkleuren er even fraai uitzien!
Omdat de naaldmatrix printer een regeldrukker is, zal het duidelijk zijn dat kleuren op het papier verschijnen door de printkop vier keer over een en dezelfde regel te sturen. Bij iedere doorgang wordt er een andere kleurband van het inktlint voor de naalden geplaatst.

Nadelen

Naaldmatrix printers met kleur hebben in feite niets dan nadelen. Jaren lang was dit principe echter het enige waarmee men op een betaalbare manier kleur op papier kon krijgen. Nu er inktjet en laserprinters op de markt zijn die hetzelfde veel beter, goedkoper en sneller kunnen, is de rol van de kleuren naaldmatrix printer volledig uitgespeeld.
Het grootste probleem met gekleurde inktlinten is dat de inktvoorraad in de vier kleurbanden niet even snel en niet gelijkmatig opraakt. Heeft men bijvoorbeeld iets geprint waar een heleboel rood in zit, dan zal het rode gedeelte van het inktlint plaatselijk minder kleur bevatten. Drukt men nadien een gekleurde illustratie af, waarbij gebruik wordt gemaakt van mengkleuren, dan kan het gebeuren dat een oranje vlak op de ene plaats een heel andere tint heeft dan op een andere plaats.

Monochrome thermo-direct printers

Principe

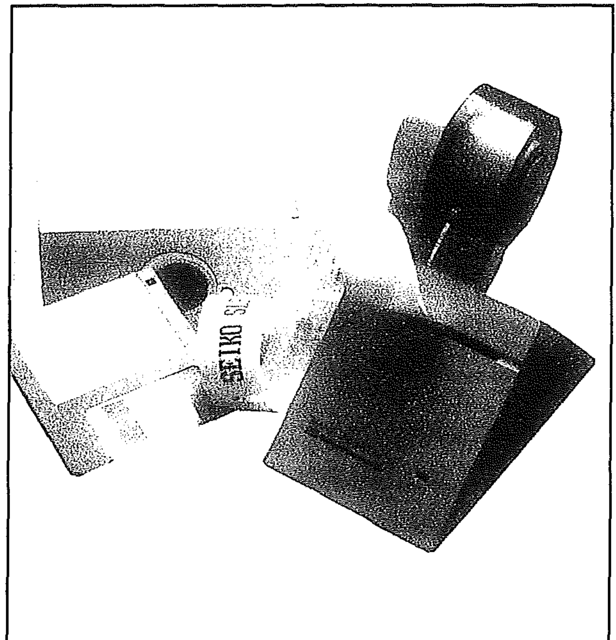
Het fundamentele werkingsprincipe van een thermo-direct printer wijkt in wezen niet erg veel af van dat van de naaldmatrix collega. Ook nu is er een printkop, die door middel van een geleidingsmechanisme van links naar recht of van rechts naar links over het papier wordt verplaatst. Ook nu wordt het papier door middel van een rubber rol regel na regel verplaatst. De printkop bestaat nu echter uit een aantal zeer kleine metalen contactvlakjes. Onder ieder vlakje zit een klein elektrisch verwarmingselementje. De thermo-direct printers werken alleen als er speciaal chemisch geprepareerd papier wordt gebruikt. Aan één zijde van het vel is een coating aangebracht met een stof, die onder normale temperatuursomstandigheden wit is,

maar zwart verkleurt als de temperatuur stijgt tot ongeveer 90 °C. De thermische printkop glijdt over het papier. Als een dot geprint moet worden, wordt het betreffende verwarmingselementje even gestuurd. Het contactvlakje wordt heet, de hitte straalt af op het papier en het oppervlak onder het contactvlakje wordt zwart.

Klassen
Printers die werken volgens het thermodirect principe behoren tot de klasse van de regelprinters. Bovendien zijn het non-impact printers.

Voor- en nadelen
Het thermo-direct principe heeft als voordeel dat de printer absoluut geruisloos werkt. Er zijn geen bewegende onderdelen, behalve dan natuurlijk de twee motoren die zorgen voor het papier- en koptransport. Maar dat kan tegenwoordig volledig in stilte gaan. De levensduur van een dergelijk apparaat is dan ook erg hoog, waarbij er bovendien nauwelijks onderhoud noodzakelijk is. Tweede groot voordeel is dat de geprinte dots diepzwart zijn, hetgeen mooi contrasteert met de witte laag hittegevoelige stof op het papier.
Het systeem heeft als groot nadeel dat er speciaal vrij duur thermisch papier voor nodig is. Bovendien is de afdruk niet stabiel. Na enige jaren verkleurt ook het niet verhitte deel van het papier, waardoor het contrast van de afdruk gaat dalen.

Toepassingen
Printers die volgens het thermo-direct procédé werken, worden tegenwoordig zelden gebruikt als uitvoerapparaat van een computer. Maar natuurlijk heeft het principe een grote verspreiding gekregen door de popularisatie van de fax. De meeste goedkope faxapparaten werken met een thermo-direct uitvoer, omdat men er van uitgaat dat faxberichten de tand des tijds niet hoeven te weerstaan en gerust mogen verkleuren.
Daarnaast wordt het procédé toegepast in speciale printers. Een van de bekendste voorbeelden zijn de in figuur 3/19.11-8 voorgestelde labelprinters van Seiko Instruments. Deze bedrukken speciaal chemisch geprepareerde zelfklevende etiketten van de rol met een thermo-direct printkop.

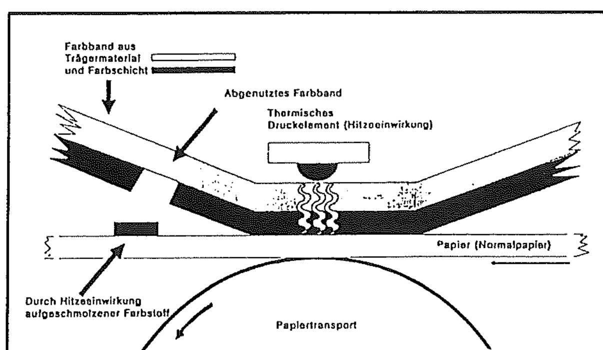


Figuur 3/19.11-8: De "smart label printer" van Seiko werkt volgens het thermo-direct principe.

Monochrome thermo-transfer printers

Principe
Het principe van deze thermische drukkers is in grote lijnen te vergelijken met

dit van de thermo-direct printers. Zij werken met hetzelfde soort printkop. Zoals uit figuur 3/19.11-9 blijkt, zit er nu echter tussen de printkop en het papier een speciaal inktlint. Een mechanisch veersysteem zorgt ervoor dat het lint zacht tegen het papier wordt aangedrukt. Het lint bestaat uit een kunststof drager, waarop een speciaal soort kleurstof is aangebracht. Deze kleurstof op wasbasis heeft een laag smeltpunt. Als een van de contactplaatjes van de printkop wordt opgewarmd, zal plaatselijk een beetje van deze was smelten. De vloeibare was wordt overgedragen op het papier en vormt daar een klein zwart puntje. Het koude papier zorgt er immers voor dat de was onmiddellijk weer stolt.



Figuur 3/19.11-9: De werking van een thermo-transfer printer.

Papiersoorten

Uit deze beschrijving van de werking volgt het grote voordeel van het systeem: men kan gewoon papier gebruiken! Toch blijkt in de praktijk dat een afdruk op normaal copieerpapier niet die hoge kwaliteit heeft, waartoe het systeem in staat is.
Normaal papier is namelijk vezelachtig en het gevolg is dat de gesmolten was gedeeltelijk in het papier smelt. Het resultaat is dat de was iets uitgesmeerd wordt, waardoor de dots geen scherpe omrandig hebben, meer min of meer de vezelstructuur van het papier volgen. Doordat een gedeelte van de was in het papier dringt is de zwarting niet optimaal.
Wil men de allerhoogste kwaliteit uit een dergelijke printer halen, dan zal men gebruik moeten maken van speciaal geprepareerd papier. Dit papier heeft een felwitte glanzende bovenlaag, die als eigenschap heeft dat de gesmolten zwarte was er op blijkt liggen, maar er zich wel uitstekend aan hecht. Het gevolg is een diepzwarte afdruk op helwit papier, met bovendien zeer scherpe omrandigen van de individuele dots. Gebruikt men dit vrij dure papier, dan kan de afdrukkwaliteit van een thermo-transfer printer zich zonder meer meten met deze van de allerbeste laserprinters.

Klassen

Het zal wel zonder nadere toelichting duidelijk zij dat ook thermo-transfer printers regelprinters zijn die werken volgens het non-impact systeem. Ook met deze printers kan men geen doorslagen maken.

De printkop

De constructie van de printkop van een thermo-transfer printer is veel en veel eenvoudiger dan deze van naaldmatrix printers. Tegenwoordig kan een thermische printkop zelfs volledig in dikke-film technologie worden uitgevoerd. Dat betekent dat alle onderdelen via optisch-chemische weg op een ceramische drager worden geëtst en/of opgedampt. Een moderne thermische printkop is dan ook maar enige mm dik en heeft erg weinig thermisch verlies. Dat betekent dat de opgewekte warmte vrijwel volledig daar terecht komt waar het de bedoeling is, namelijk bij de dunne contactplaatjes die tegen het inktlint aanwrijven. Er moet dus heel weinig elektrisch vermogen in de kop gepompt

19.11 Principes en werking van printers

worden, hetgeen als groot voordeel heeft dat de printer gemakkelijk uit batterijen gevoed kan worden.

De resolutie

Het dikke-film procédé heeft als voordeel dat het erg goedkoop mogelijk is vele elementjes op een klein oppervlak onder te brengen. Het is mogelijk tot 32 elementjes dual in-line staggered op te stellen, waardoor een thermo-transfer printer een resolutie haalt die zonder meer te vergelijken is met, of zelfs beter is dan, deze van een goedkope laserprinter. Zelfs de goedkoopste thermo-transfer printers beschikken over een resolutie van 300 dpi en de betere modellen halen tegenwoordig 400 dpi.
In figuur 3/19.11-10 is als bewijs van de goede printerkwaliteit het vier maal vergrote printbeeld van een thermo-transfer printer voorgesteld.

Voor- en nadelen

Thermo-transfer printers zijn heel handige apparaatjes, die uitstekende drukresultaten kunnen opleveren. Bovendien kunnen zij zeer compact gebouwd worden. Omdat er in de printkop geen ingewikkelde mechanische miniatuurconstructies verzonnen moeten worden, kan een dergelijke kop zeer klein worden gemaakt. Thermo-transfer printers zijn tegenwoordig nauwelijks groter dan het velletje A4 dat zij moeten bedrukken en niet dikker dan enige centimeter. Men is er reeds in geslaagd dergelijke printers te integreren in laptop's, zodat het mogelijk is onderweg snel even een briefje uit te printen.
Het grote nadeel van het thermo-transfer principe is dat de inktlinten slechts éénmaal bruikbaar zijn. Daar waar de kleurstof is gesmolten ontstaat immers een onderbreking in de kleurlaag. Dergelijke linten zijn dus te vergelijken met de oude carbonlinten van de schrijfmachines. Ook die konden maar eenmalig gebruikt worden. Wie dus erg veel te printen heeft, doet er verstandig aan de inktrollen per gros in te kopen!
Een tweede nadeel is dat de afdruk tamelijk gevoelig is voor beschadigingen. Dit is zeker het geval als men gebruik maakt van het speciale glanzende papier, waarbij de wasdotjes op het papier blijven liggen. Krast men met een scherp voorwerp over de afdruk, dan laten de wasdotjes los en ontstaat er een lelijke witte kras op de afdruk.

The quick
The quick
The quick
The quick
The quick

Figuur 3/19.11-10: Het vier maal vergrote printbeeld van een thermo-transfer printer.

Beveiligen van vertrouwelijke gegevens

Het lijkt misschien raar om het in een hoofdstuk over de technologie van printers te hebben over beveiligen van ver-

trouwelijke gegevens. Toch is dit een punt waar iedereen die thermo-transfer printers gebruikt goed rekening mee moet houden! Zoals beschreven kan het inktlint maar eenmalig gebruikt worden, omdat er bij het printen inkt van het inktlint naar het papier wordt overgebracht. Het gevolg is dat op het inktlint een spiegelbeeld ontstaat van alle letters die zijn afgedrukt! Wie een gebruikt thermisch lint uit de prullenbak vist en er mee voor een spiegel gaat staan, kan zonder problemen alle brieven, die de afgelopen dagen of weken geschreven zijn, lezen. Hetzelfde geldt uiteraard voor het lint dat nog in de printer zit.

Thermo-transfer printers als paginaprinters

Tegenwoordig worden thermo-transfer printers ook aangeboden als paginaprinters. Deze werken met een statische (dus niet bewegende) printkop, die net zo breed is als het te bedrukken papier. Op deze constructie zijn duizenden minuscule verwarmingselementjes naast elkaar aangebracht. Voor een A4-printer met een resolutie van 300 dpi heeft men niet minder dan 2.560 elementjes nodig. Bij hogere resoluties werkt men met enige rijen elementen onder elkaar, waarbij de positie van de elementen op iedere rij iets verschoven is. Dit is dus te vergelijken met de dual-in-line staggered technologie van de matrix printkoppen. Ook het inkt-"lint" is net zo breed als het papier, zodat men in feite beter kan spreken van een inktrol. Het drukken begint pas op het moment dat de printer alle gegevens die op de pagina moeten komen van de computer heeft ontvangen. Nadien worden papier en inktrol synchroon aangedreven en worden de te printen gegevens vertaald naar een reeks parallelle besturingspulsen voor de duizenden naast elkaar gelegen verwarmingselementen. Op die manier wordt dot-rij na dot-rij over de volle breedte van het papier gedrukt.

Nadeel van de paginaprinter

Een groot nadeel van de thermo-transfer printer die werkt met een inktrol is dat het verbruik van deze rol volledig onafhankelijk is van de hoeveelheid gegevens die op een pagina moeten worden gedrukt. Voor het bedrukken van één A4-vel wordt een lengte van de inktrol verbruikt, die precies gelijk is aan de lengte van het papiervel. Dat wordt veroorzaakt door het synchrone transport van papier en inktrol. Zelfs als op het te bedrukken vel maar één regeltje tekst staat, wordt de inktrol toch over de lengte van een vel A4 verder getransporteerd.

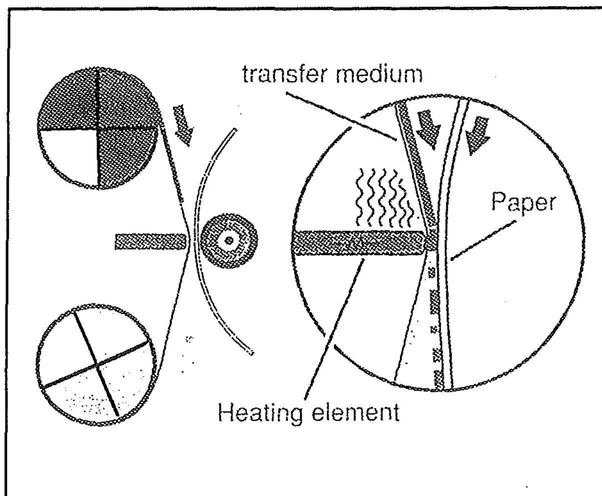
Kleuren thermo-transfer printers

Principe

Het principe van thermo-transfer printers die met een inktrol werken, is om te zetten naar een kleursysteem. De inktrol bestaat nu uit een opeenvolging van drie of vier folies met verschillend gekleurde stoffen, die allemaal even groot zijn als het te bedrukken papier, zie figuur 3/19.11-11. Deze rol wordt weer door een verwarmings-array, bestaande uit duizenden verwarmingselementjes, tegen het papier gedrukt. Het volledige kleurenbeeld wordt in drie of vier arbeidsgangen samengesteld. Eerst wordt de eerste kleurfolie in de juiste positie gebracht. Bij de eerste drukgang wordt het papier in deze ene kleur bedrukt. Nadien wordt het papier

19.11 Principes en werking van printers

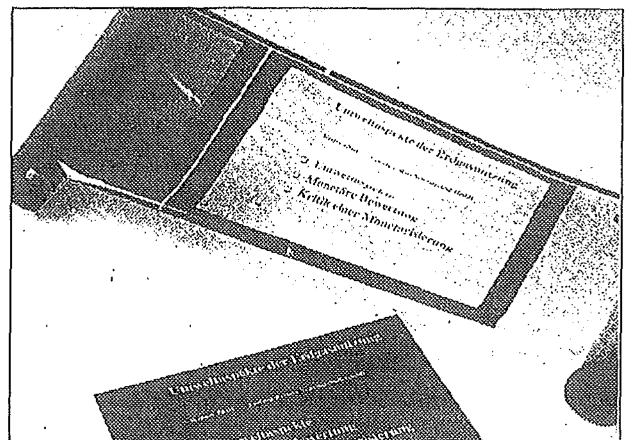
terug naar de uitgangspositie getransporteerd, de tweede kleurfolie in positie gebracht en wordt de tweede kleur aangebracht. Dit gaat zo verder tot alle drie of vier kleuren op het papier zijn aangebracht.
Het zal duidelijk zijn dat er bij dergelijke kleurenprinters heel hoge eisen worden gesteld aan de nauwkeurigheid van het papiertransport. Het papier gaat drie of vier keer heen en weer in het mechanisme. Zit er speling in dit transport, dan zullen de drie of vier kleurenbeelden elkaar niet precies dekken en ontstaan zeer storende kleurzwemen aan de randen van de kleurovergangen. Dat probleem mag niet onderschat worden en de constructieve oplossing ervan bepaalt in hoge mate de kwaliteit van de printer. De onnauwkeurigheid van het papiertransport moet bij moderne 300 dpi thermo-transfer printers kleine zijn dan 100 µm!



Figuur 3/19.11-11: Het principe van de thermo-transfer printer met kleur.

Voor- en nadelen

Ten opzichte van de nog te bespreken inkjet printers hebben de thermo-transfer printers als groot voordeel dat de intensiteit van de gedrukte kleuren constant blijft. Voor iedere afdruk wordt er immers een nieuw, volledig gaaf stuk van de inktrol gebruikt, zodat dergelijke printers geen last hebben van leeg rakende inktpatronen. Dit voordeel is niet zo wezenlijk bij zwartdruk, maar wel bij kleurendruk. Hier, immers, geeft iedere afwijkende dekking van een van de drie basiskleuren een zeer vertekende mengkleur. Thermo-transfer drukkers zijn dan ook ideaal voor het produceren van kwalitatief hoogwaardige kleurenprints, waar veel met mengkleuren wordt gewerkt. Te denken valt hierbij aan proeven van pagina's die in full-colour worden opgemaakt en de klant een uitstekend idee geven van hoe het drukwerk er uiteindelijk zal komen uit te zien.



Figuur 3/19.11-12: Een stukje van de verbruikte inktrol na het bedrukken van één vel papier.

Wel moet opgemerkt worden dat de nog te bespreken kleurstof sublimatie printers op dit gebied nog beter presteren. Nadeel is wel dat de afdrukkosten erg hoog zijn, dit als gevolg van de slordige en oneconomische manier, waarop een thermo-transfer printer met de inktrol omgaat. Iedere A4-afdruk in vierkleurendruk kost

19.11 Principes en werking van printers

niet minder dan 1,18 meter van de inktrol! In figuur 3/19.11-12 kan men heel mooi zien hoe de inktrol er uit ziet nadat de printer de onderste afdruk heeft gemaakt: enige witte tekstregels tegen een verlopende, kleurige achtergrond.

Monochrome inkjet printers

Inleiding

De ontwikkeling van het inktjet principe is een hoogstandje van de moderne technologie. Voor de fabricage van de printkoppen worden micron-technologieën toegepast, die voorheen alleen bekend waren van de fabricage van geïntegreerde schakelingen. Om het gedrag van vloeistoffen die stromen door kanaaltjes van minder dan een duizendste millimeter diameter te begrijpen moest fundamentele fysische research verricht worden op het gebied van vloeistofdynamica, waaronder de eigenschappen van laminaire en turbulente stromingen. Om de inkt de noodzakelijke vaak tegenstrijdige eigenschappen te geven werden alle registers van de moderne chemie opengetrokken. Ondanks dat is het principe niet nieuw. De eerste primitieve printer die werkte met een gecontroleerde vloeistofstroom werd reeds in 1952 ontwikkeld door Siemens. Deze "Mingograf" werd gebruikt voor het registreren van de elektrische hersenactiviteiten op een papierband.

Principe

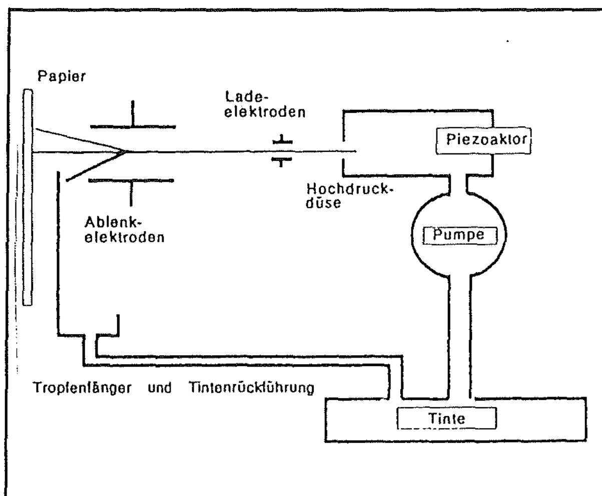
Een inkjet printer spuit kleine druppeltjes inkt op het papier. De printkop is voorzien van een groot aantal uiterst kleine gaatjes, de zogenoemde "nozzles" of spuitmondjes. De eerste apparaten die op de markt kwamen waren erg duur en gaven bovendien nogal wat problemen, onder meer doordat de uiterst fijne nozzles om de haverklap verstopten. Het was Hewlett Packard die met zijn serie DeskJet-printers een ware technologische revolutie veroorzaakte. Met deze serie kwam een reeks goedkope en kwalitatief hoogstaande printers op de markt, waarvan de printresultaten zich in weinig onderscheiden van de (toen) veel duurdere laserprinters.
De printkop van een inkjet printer is een waar wondertje van moderne technologie. Op een minuscuul oppervlak van een paar vierkante millimeter zijn tussen 48 en 64 spuitmondjes aanwezig, waarin bepaalde mechanismen er voor zorgen dat ongeveer 100 picoliter inkt (is gelijk aan 0,000.000.000.1 liter!) uit een reservoirtje wordt opgepompt en op de juiste momenten met een snelheid van 40 km/h naar het papier wordt gespoten. Hiervoor zijn in de loop der tijden drie verschillende systemen ontwikkeld, namelijk:
- het continuous drop principe;
- het bubblejet principe;
- het piëzo-elektrisch principe.

Continuous drop principe

Het allereerste principe dat werd ontwikkeld, werkte met een continue inktstroom. Vandaar de naam "continuous drop". Deze continue inktstroom moest er voor zorgen dat het grote probleem van het principe, namelijk het dichtslibben van de nozzles, nooit kon voorkomen. Het principe is geschetst in figuur 3/19.11-13. Uit het inktreservoir "Tinte" wordt via een pomp "Pumpe" en een hoge druk reservoir "Hochdruckdüse" een uiterst dunne inktstraal naar het papier gespoten. Deze straal wordt door twee oplaadelektroden "Ladeelektroden" geleid, die de molecu-

19.11 Principes en werking van printers

len van de inkt elektrisch opladen. Voordat de inktstraal op het papier terecht komt, wordt zij eerst door twee afbuigelektroden "Ablenkelektroden" gevoerd.

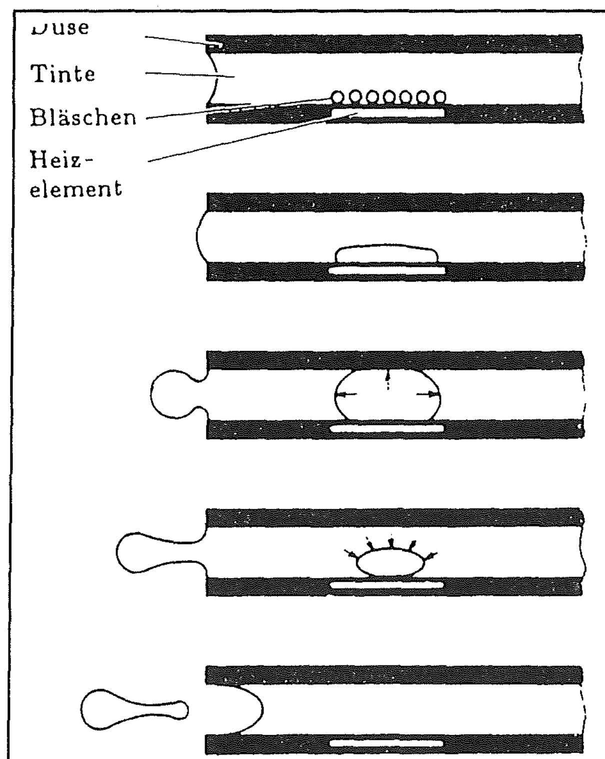


Figuur 3/19.11-13: Het principe van het continuous drop systeem.

Tussen deze twee elektroden wordt een positieve of negatieve gelijkspanning gezet. Het elektrisch veld dat hierdoor ontstaat, zal de geladen inktdruppeltjes ofwel afbuigen in de richting van het papier, ofwel afbuigen naar een opvangmechanisme "Tropfenfänger". Via een dunne leiding wordt de opgevangen inkt dan weer terug gevoerd naar het reservoir. Moderne inktsoorten hebben de eigenschap dat zij alleen drogen als zij in contact komen met de zuurstof uit de lucht. De kans dat zij in het printmechanisme uitdrogen is daardoor uiterst klein. De vrij ingewikkelde constructie die voor het continue principe nodig is, is bij deze moderne inktsoorten dan ook niet meer noodzakelijk. Vandaar dat het "continuous drop"-principe verlaten is en vervangen door het zogenoemde "drop on demand"-principe. Hierbij wordt alleen een druppeltje inkt geproduceerd op het moment dat het echt noodzakelijk is, dus op het moment dat er inkt op het papier moet komen.

Bubblejet principe

Bij het eerste "drop on demand"-principe wordt gebruik gemaakt van het feit dat stoffen die verdampen uitzetten. In iedere inkttoevoerleiding naar een nozzle zit, zie figuur 3/19.11-14, een klein verwarmingselementje "Heizelement". In de meeste gevallen heeft dit een weerstand van ongeveer 30 Ω. Stuurt men een stroomstootje door dit element, dan warmt dit op en zal de warmte overdragen naar de inktmoleculen in de omgeving.



Figuur 3/19.11-14: Het principe van de bubblejet printkoppen.

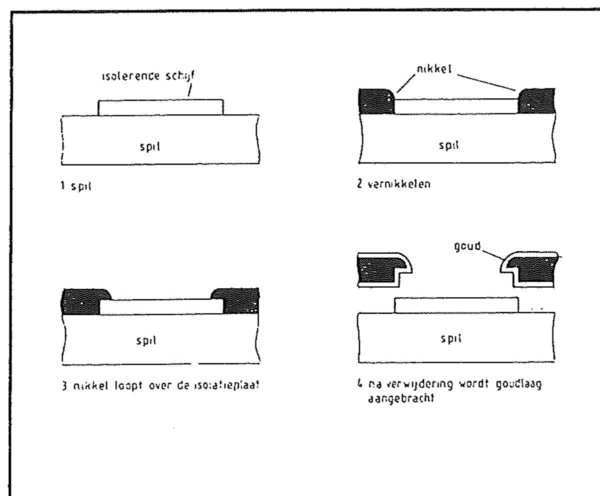
De inkt verdampt, waardoor het volume met duizenden procenten toeneemt. Hierdoor wordt een bepaalde hoeveelheid inkt met grote kracht uit de nozzle gedrukt en naar het papier gespoten. Als

19.11 Principes en werking van printers

de stroompuls door het weerstandje wegvalt zal het dampbelletje weer condenseren, de druk in het buisje valt weg en de inkt wordt weer naar binnen gezogen.
Het principe lijkt eenvoudig, maar voordat de eerste printkop, die volgens dit principe werkt, op de markt verscheen is er een enorme hoeveelheid research gepleegd! Het is Hewlett Packard die hier baanbrekend werk heeft verricht.
Een groot probleem is de snelheid, waarmee de inktdruppeltjes kunnen worden weggeschoten. Deze snelheid bepaalt de uiteindelijke printsnelheid en wordt begrensd door de snelheid waarmee men de verwarmingselementjes op temperatuur kan brengen en nadien weer kan laten afkoelen. Om de noodzakelijke frequentie van 3.000 druppeltjes per nozzle per seconde te halen, worden er zeer hoge eisen gesteld aan de constructie van het gehele mechanisme.
Bij de printkop van de oude PaintJet's van Hewlett Packard wordt gewerkt met dertig nozzles. De verwarmingselementen liggen op een onderlinge afstand van slechts 200 µm en de lengte van een spuitmond bedraagt slechts 50 µm.

De constructie van de printkop

Dergelijke kleine afmetingen kunnen alleen gemaakt worden door gebruik te maken van dunne film technieken. Het productieproces is geschetst in figuur 3/19.11-15.
Basis is een plaatje roestvrij staal, de zogenoemde spil, dat voorzien wordt van een elektrisch niet geleidende film. De vorm van de film bepaalt de uiteindelijke vorm van de plaat. De plaat wordt in een vernikkeltank geplaatst en voorzien van een negatieve spanning. Op de plaatsen waar de film is aangebracht, zal geen vernikkeling plaats vinden.

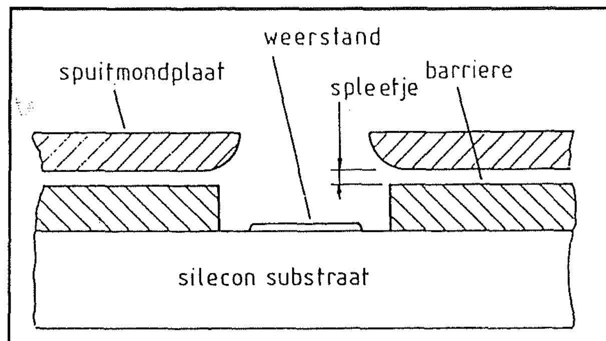


Figuur 3/19.11-15: De constructie van de nikkel spuitmondplaat waarop de nozzles worden aangebracht.

Door het vernikkelproces iets langer te laten duren dan noodzakelijk, zal de opgebrachte nikkellaag iets over de filmlaag aangroeien. Hierdoor worden mooie afgeronde materiaalovergangen verkregen, die er voor zorgen dat de inkt zonder turbulente stromingen door het systeem kan vloeien. Nadien wordt de uiterst dunne nikkellaag gescheiden van de roestvrij stalen drager en wordt het nikkel verguld. Op deze manier wordt de spuitmondplaat gemaakt.
De verwarmingselementen worden gevormd op een plaatje silicium als basis. Ieder element heeft een dikte van slechts 0,1 µm, maar kan toch temperaturen van 400 °C genereren. De warmtepuls mag slechts 5 µs lang zijn om de beoogde printsnelheid te halen. Op het plaatje silicium wordt eerst een dunne laag aluminium opgedampt. Daarboven op komt een laagje siliciumoxide, dat dient voor de warmteverdeling, waarop de verwarmingselementjes worden aangebracht via een opdampproces. Tot slot wordt de silicium basisplaat met de verwarmingselementjes verenigd met de nikkelen spuitmond-

19.11 Principes en werking van printers

plaat. Tussen beide platen wordt een barrière van slechts een paar µm aangebracht, waardoor de inkt stroomt (zie figuur 3/19.11-16).

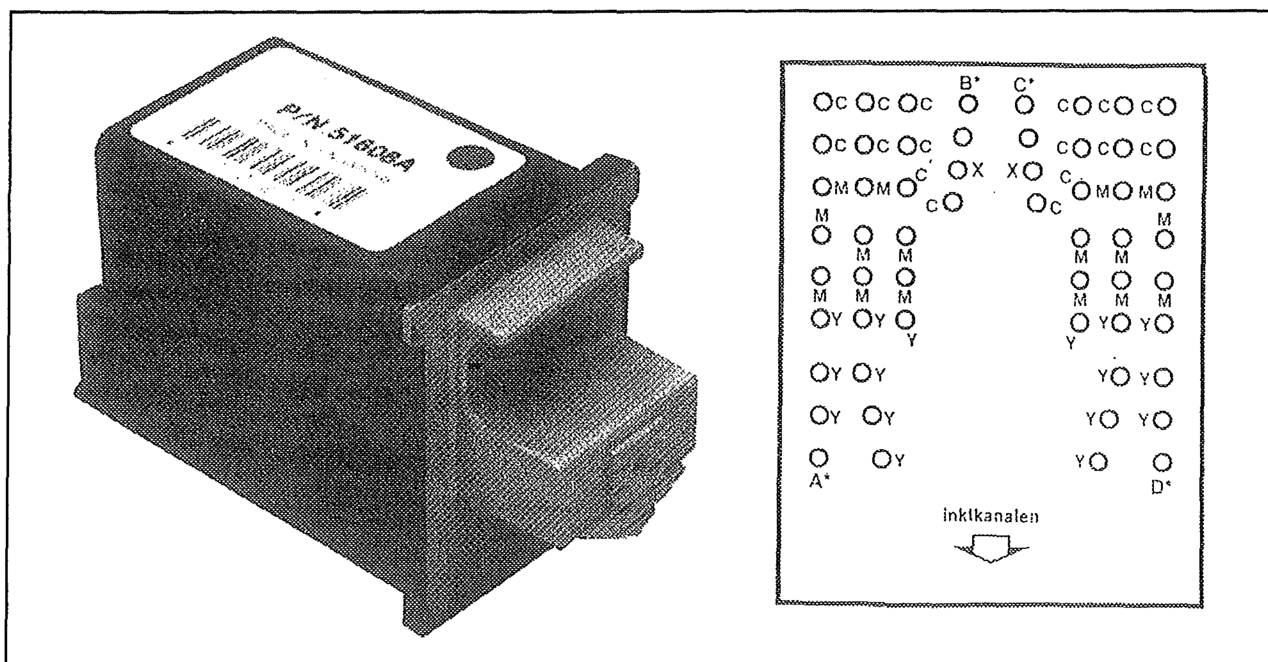


Figuur 3/19.11-16: De silicium basisplaat met de verwarmingselementen wordt verenigd met de nikkelen spuitmondplaat.

De gehele constructie, overigens maar een paar vierkante millimeter groot, wordt nu aangebracht op een flexibele gedrukte bedrading. Hier worden de contacten naar de verwarmingsplaatjes verzorgd en aan de andere kant de contacten naar het mechanisme van de printer. Deze flexibele gedrukte bedrading wordt nu op het inktpatroon bevestigd.

Het standaard inktpatroom van de HP-printers wordt voorgesteld in figuur 3/19.11-17, waarbij tevens een indruk wordt gegeven van het aantal contactpuntjes dat noodzakelijk is om alle nozzles aan te sturen.

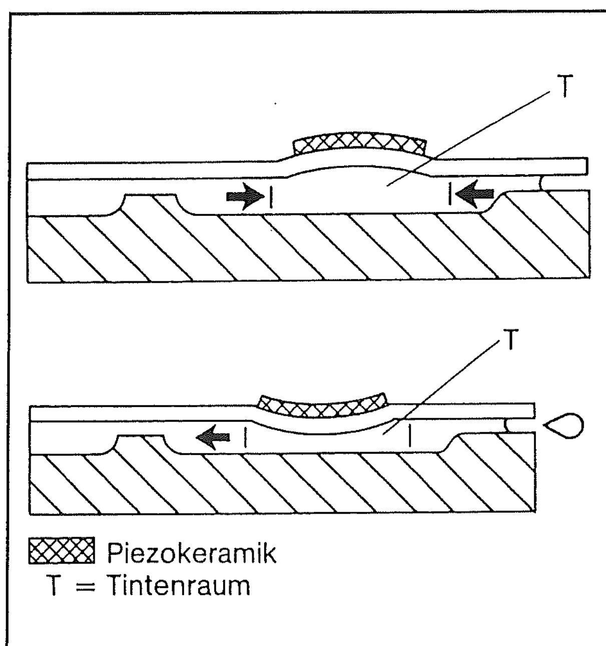
Als het patroon in de printer wordt bevestigd, maken de vergulde contactpuntjes op het patroon contact met even zoveel connectorcontacten in de behuizing van de printkop. De zwaartekracht en de capillaire zuigkracht van de barrières zorgen ervoor dat de dun vloeibare inkt naar de spuitmondjes vloeit.



Figuur 3/19.11-17: Een inktpatroon voor een inktjet printer, met rechts de gestandaardiseerde codering van de contactpuntjes, waarmee de 48 nozzles worden aangestuurd. De codes C, M en Y verwijzen naar de drie basiskleuren.

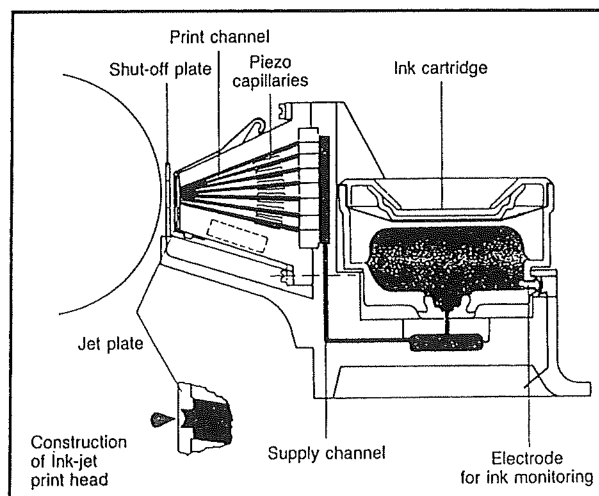
Het piëzo-elektrisch principe

Bij dit principe, dat door Epson en Canon wordt toegepast, wordt gebruik gemaakt van een ander fysisch verschijnsel, namelijk de piëzo-elektriciteit. Bepaalde ceramische materialen zullen, als zij onder elektrische spanning worden gezet, vervormen. Dit piëzo-elektrisch effect wordt onder andere toegepast in sommige microfoons en in kristallen. Als men van een dergelijk materiaal een uiterst dun plaatje maakt en een kleine gelijkspanning over het plaatje zet, dan zal het plaatje krom trekken. Door deze beweging kan men een klein inktdruppeltje uit een nozzle schieten. De constructie van een dergelijke piëzo-elektrische nozzle is getekend in figuur 3/19.11-18.



Figuur 3/19.11-18: De schematische constructie van een piëzo-elektrische nozzle.

Ook nu bestaat de constructie uit een basisplaat, waarboven op een zeer kleine afstand een spuitmondplaatje wordt bevestigd. Dit plaatje bezit nu echter een bepaalde mate van flexibiliteit. Op dit spuitmondplaatje zijn de zeer kleine piëzo-ceramische plaatjes (gearceerd weergegeven) aangebracht. In rusttoestand trekken deze plaatjes de flexibele spuitmondplaat plaatselijk naar boven. Er wordt nu een kleine ruimte gevormd, die zich vol zuigt met inkt. Zet men het piëzoceramische plaatje onder spanning, dan zal dit vervormen, waardoor er druk op de inkt wordt uitgeoefend. Het ligt voor de hand dat deze de weg van de minste weerstand zoekt en dus de grootste hoeveelheid inkt naar rechts een uitweg zoekt. Op deze manier wordt een heel klein druppeltje inkt naar het papier gespoten. In figuur 3/19.11-19 is de constructie van een volledige printkop getekend. De inkt uit de inkt cartridge wordt via een kanaaltje naar de nozzles gevoerd en wordt via de diverse printkanaaltjes op het papier gespoten.



Figuur 3/19.11-19: De constructie van een printkop op piëzo-ceramische basis.

Fundamenteel verschil

Er bestaat een zeer fundamenteel verschil tussen printkoppen op bubble-jet en op piëzo-elektrische basis. Bij de eerstge-

19.11 Principes en werking van printers

noemde soort vormt het printmechanisme een onlosmakelijk deel uit van het inktreservoir. Als dit leeg is, moet ook het nozzle mechanisme vervangen worden. Bij de piëzo-elektrische koppen staat het inktreservoir los van het mechanisme. De kop, met de nozzles en ceramische plaatjes gaat veel langer mee, alleen de inkt cartridge wordt vervangen.

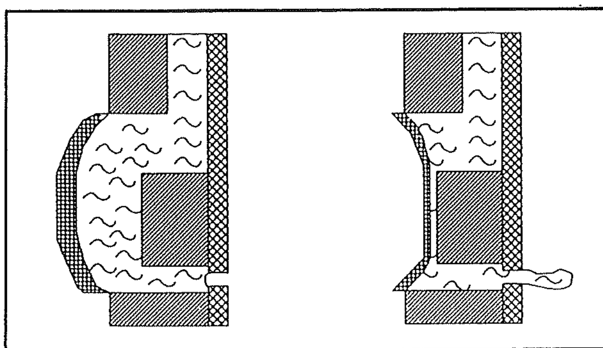
Klassen
Het zal wel duidelijk zijn dat inktjet printers non-impact printers zijn. Met als groot voordeel dat het printen zélf volledig geruisloos werkt en als groot nadeel dat men geen doorslagen kan printen. Wat betreft regel- of paginaprinters kan gesteld worden dat het principe zich voor beide systemen leent. In de meeste gevallen is het de software die bij de printer geleverd wordt die bepaalt hoe de printer moet werken. Ontvangt de printer een zuiver tekstbestand, dan zal de software de printer in regelmodus schakelen en regel na regel printen. Ontdekt de software dat ingewikkelder zaken geprint moeten worden, dan zal de printer als paginaprinter gestuurd worden.

Voor- en nadelen
Inktjet printers hebben als groot voordeel dat zij tegenwoordig heel goedkoop aangeboden worden, terwijl de kwaliteit van de afdruk even goed is als deze van duurdere laserprinters. Ook de resolutie kan zich meten met deze van de beste laserprinters. Resoluties van 360 dpi zijn tegenwoordig standaard, terwijl diverse inktjet printers beschikbaar zijn met een resolutie van 720 dpi. Met monochrome inktjet printers kan men zonder enig bezwaar gewoon goedkoop papier bedrukken, terwijl voor het betere werk speciaal glanzend papier leverbaar is. Voor printers die alleen zwart drukken heeft het gebruik van dit dure papier echter weinig zin. Het aloude probleem van het indrogen van de inkt in de nozzles doet zich bij de moderne cartridges niet meer voor. De printers zijn voorzien van een intelligent bestuurd reinigingssyteem, waarbij de spuitopeningen worden schoon gemaakt en dit niet alleen voor iedere printopdracht, maar ook regelmatig tijdens het printen! In de ruststand wordt de printkop naar een bepaalde positie gestuurd, waarbij de spuitopeningen op een stukje rubber worden geperst. Hierdoor worden de nozzles luchtdicht afgesloten.

Eigenschappen
Monochrome inktjet printers hebben over het algemeen de onderstaande eigenschappen:
- levensduur mechanisme: 100 miljoen tekens;
- levensduur inktpatroon: een miljoen tekens;
- snelheid: 100 tot 200 cps in sneldruk, 30 tot 100 cps in schoonschrift.

Hoge snelheid printers
Voor professionele toepassingen worden monochrome inktjet printers aangeboden die de ongelofelijke hoeveelheid van 20.000 inktdruppeltjes per seconde naar het papier kunnen spuiten. Dit wordt verwezenlijkt door de nozzles uit te voeren volgens de "edge shooter"-technologie. Hierbij bestaat er tussen de invoer en de uitvoer van de inkt een hoek van negentig graden (zie figuur 3/19.11-20). Uit de constructietekening blijkt dat het uitstootkanaal van de inkt nu zeer kort is, in de meeste gevallen niet meer dan enige tientallen μm. Wie figuur 3/19.11-20 vergelijkt met figuur 3/19.11-19 zal onmiddellijk begrijpen dat door dit zeer korte uit-

voerkanaal een zeer aanzienlijke verhoging van de werkingsfrequentie van de nozzle zal optreden.



Figuur 3/19.11-20: De constructie en werking van een nozzle volgens het "edge shooter"-principe.

Kleuren inktjet printers

Eenvoudig uit te breiden

Het zal duidelijk zijn dat het principe van de inktjet printer zich uitstekend leent voor kleurendruk. Men moet immers alleen het aantal printkoppen uitbreiden tot vier, deze vier koppen keurig naast elkaar opstellen en iedere printkop vullen met de geëigende kleur. Dat is dan ook precies wat in de praktijk gebeurt!

Bij de goedkope typen kleurenprinters wordt een tweede inktkop naast de zwarte gemonteerd, waarin een speciale inktcartridge past die drie kleurcompartimenten heeft. De drie spuitmonden hebben echter veel minder nozzles dan bij de zwarte cartridge, waaruit automatisch volgt dat de maximale resolutie voor kleur lager is dan deze bij zwart/wit. Het nadeel van dit soort printers is dat het inktverbruik zeer oneconomisch is, omdat het zelden zal voorkomen dat men even veel rode, gele en blauwe inkt verbruikt. Als een van de compartimenten leeg is, moet men de kleurencartridge in zijn geheel vervangen.

Bij de professionele, duurdere modellen gebruikt met vier identieke printkoppen en cartridges, ieder voor een kleur. Het zal duidelijk zijn dat de resolutie voor kleur nu even hoog kan zijn als deze voor zwartdruk, terwijl geen inkt verloren gaat. Maar door de ingewikkelde constructie zijn deze printers uiteraard veel duurder.

Voor- en nadelen

Inktjet kleurenprinters hebben als grootste voordeel dat zij de goedkoopste kleurenprinters zijn, die acceptabele resultaten leveren. Toch valt er nog wel het een en ander aan te merken op de afdrukkwaliteit. Afdrukken op normaal papier zijn in het algemeen onacceptabel. Het papier zuigt de kleuren op, waardoor de afdruk een onverzadigde indruk achter laat. Heeft het papier een vezelstructuur, dan zal de inkt voornamelijk langs de vezels in het papier dringen, waardoor de kleuren in elkaar overvloeien en een onscherpe indruk bestaat. Bij het printen op te glad papier bestaat het gevaar dat de kleuren te langzaam indrogen, waardoor de natte inkten zich kunnen gaan mengen en allerlei onnatuurlijke kleurverlopen ontstaan. Dit laatste is zeer zeker het geval als men grote naast elkaar liggende vlakken in verschillende kleuren moet bedrukken. De meeste fabrikanten leveren echter speciaal papier, dat er op het oog glanzend uitziet maar ruw aanvoelt. Gebruikt men dit zeer dure papier, dan kan men met kleuren inktjet printers prachtige resultaten behalen.

Het afdrukken van foto's

Voor het afdrukken van zwart/wit- en kleurenfoto's zijn de allernieuwste inktjet

printers uitstekend geschikt. Men gebruikt zeer intelligente softwarematig aansturingsalgoritmes, waardoor prachtige resultaten ontstaan. Vergeleken met de afdruk van de veel duurdere sublimatieprinters, moeten de inktjet printers op dit gebied echter toch nog hun meerdere erkennen.

Monochrome laserprinters

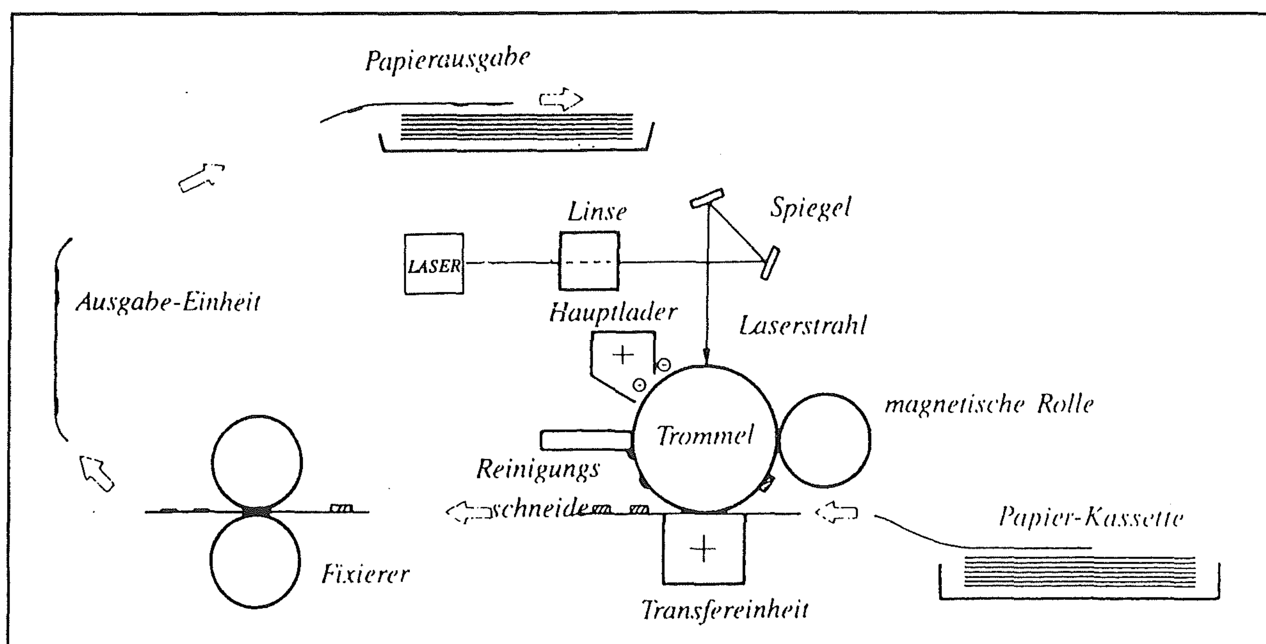
Inleiding

Wie het woord "laser" hoort, zou verwachten dat een laserprinter alles te maken heeft met een dergelijke lichtbron. Voor de meeste typen geldt dat inderdaad, maar er zijn ook printers die als laserprinter worden verkocht, maar waar in geen velden of wegen iets te ontdekken valt dat op een laser lijkt. Vandaar dat de naam laserprinter eigenlijk misleidend is. Veel beter zou men deze apparaten kunnen samenvatten onder het verzamelbegrip *"printers die werken met poedervormige inkt"*. Want dat is in feite de enige eigenschap die alle soorten gemeen hebben. In deze beschrijving wordt echter toch maar de algemeen aanvaarde term laserprinter aangehouden.

In principe werken de meeste laserprinters op dezelfde manier als een kopieerapparaat. Het basisprincipe heet "elektrostatische beeldoverdracht", ook wel eens "xerografie" genoemd, naar de fabrikant Rank Xerox, die het voor het eerst toepaste. Bij elektrostatische beeldoverdracht wordt een lichtbeeld overgedragen op papier via de tussenstap van een trommel, die elektrisch geladen wordt en inktstof aantrekt. Dat inktstof wordt "toner" genoemd. Het basisprincipe van een laserprinter is getekend in figuur 3/19.11-21. Hart van het apparaat is de trommel, ook wel "drum" genoemd. Deze is voorzien van een speciaal lichtgevoelig oppervlak, waarvoor meestal een of andere halfgeleider wordt gebruikt. Vroeger werd hiervoor het nogal giftige selenium gebruikt, tegenwoordig werkt men met amorf silicium. Naast het voordeel van niet-giftigheid is silicium bovendien veel lichtgevoeliger dan selenium. Daarnaast worden ook vaak zogenoemde "OCP"-drum's gebruikt, hetgeen de afkorting is van "Organic Photo Conductor", oftewel fotogeleider uit organisch materiaal.

De trommel draait in uurwijzerzin. Beschouw een bepaald punt op het oppervlak van de trommel. Dat punt draait eerst onder de "Reinigung" door. Hier wordt alle toner van de vorige afdruk, die nog op de trommel aanwezig is, verwijderd. De reinigingseenheid bestaat meestal uit een soort schraper of een aantal borsteltjes, die het oppervlak van de drum schoonvegen en het oude tonerpoeder afvoeren naar een afvalcontainertje. Vervolgens draait het beschouwde punt door onder de "Hauptlader". Deze unit bestaat uit een elektrostatische oplader, een zeer dunne strak gespannen metalen draad, die op een zeer hoge gelijkspanning wordt gezet. Deze draad noemt men de "primaire corona". De hoge gelijkspanning op deze draad wekt een zeer sterk elektrostatisch veld op, met als gevolg dat het oppervlak van de drum elektrisch wordt opgeladen. Vervolgens draait het punt door naar de plaats waar de echte actie plaats vindt. De drum wordt getroffen door een zeer fijne straal licht. Bij zeer fijn moet men denken aan een lichtstraaltje met een diameter van een paar duizendsten van een millimeter!

19.11 Principes en werking van printers



Figuur 3/19.11-21: Het principe van elektrofotografie.

Dat licht is afkomstig van de belichtingseenheid. Hiervoor bestaan verschillende systemen, die in een volgende paragraaf nader aan de orde komen. Bij de allereerste generatie laserprinters kan men stellen dat de dunne lichtbundel afkomstig is van een laser-eenheid. De dunne lichtstraal wordt via lenzen en beweegbare spiegels op de drum gefocusseerd. De beweegbare spiegels zorgen ervoor dat de lichtstraal uiterst snel de volledige breedte van de drum afscant. De lichtstraal wordt bovendien gelijktijdig gemoduleerd, dat wil zeggen dat de continue lichtstraal wordt omgezet in een opeenvolging van korte lichtflitsjes. Op deze manier wordt de printinformatie, die uiteraard uit logische AAN- en UIT-signalen bestaat, omgezet in lichtpulsen. Deze lichtpulsen treffen de drum. Nu gebeurt er iets vreemds. Daar waar de lichtpulsen het oppervlak van de drum treffen, zal de door de corona aangebrachte elektrische statische lading "afvloeien". Het fysische proces dat hiervoor verantwoordelijk is, is vrij ingewikkeld en er komt zelfs quantummechanica aan te pas om het verschijnsel te verklaren. In het kort komt het er op neer dat de hoog-energetische fotonen, waaruit het licht bestaat, bepaalde eigenschappen van de atomen van de lichtgevoelige laag beïnvloeden. Belangrijk is, dat de door de corona egaal opgeladen drum door de belichting een "elektrostatisch beeld" krijgt. Het beeld dat later op het papier wordt afgedrukt, is reeds op het oppervlak van de drum aanwezig, maar dan onder de vorm van elektrische ladingsverdelingen.

Beeldopbouw

De drum draait verder in uurwijzerzin, het beschouwde punt komt nu bij de magnetische rol. Deze zorgt ervoor dat het tonerpoeder zich hecht op die delen van de drum die ontladen zijn door de laserstraal. Het tonerpoeder wordt eerst opgeladen tot hetzelfde elektrisch potentiaal als de drum. De magnetische rol brengt een zeer dunne laag van dit opgeladen

poeder in de nabijheid van de drum. Daar waar de drum ontladen is, ontstaat een elektrostatisch veld tussen de ontladen delen van de drum en het elektrisch geladen tonerpoeder. Het poeder wordt aangetrokken door de ontladen delen van de drum, springt van de magnetische rol naar de drum en hecht zich op die plaatsen die ontladen zijn. In deze fase van het proces wordt het elektrostatische beeld op de drum dus omgezet in een beeld dat bestaat uit los tonerpoeder.

Hechting op papier

Ondertussen is het papier aangevoerd. Dat papier beweegt met dezelfde snelheid als de omtreksnelheid van de drum. De bedoeling is nu dat het aan de drum hechtende tonerpoeder wordt overgedragen op het papier. Ook dat gebeurt elektrostatisch. In de transfereenheid wordt het papier elektrisch opgeladen en wel met een tegengestelde polariteit als de drum. In die eenheid is dus een tweede coronadraad aanwezig, die ook op een hoge gelijkspanning staat, maar van tegengestelde polariteit als de spanning waarmee de drum werd opgeladen. Het papier wordt nu in de nabijheid van de drum gebracht. De elektrostatisch lading op het papier zorgt er nu voor dat het losse op de drum aanwezige tonerpoeder van de drum wordt losgerukt en op het papier terecht komt.

Het tonerpoeder is nu weliswaar op het papier aanwezig, maar nog steeds onder de vorm van een losse laag, die er zo vanaf gestreken kan worden. In de laatste fase van het proces wordt dit losse poeder op het papier gefixeerd. Dat gebeurt in de fixeereenheid. Het papier wordt tussen twee rubber rollen gevoerd, die door middel van een lamp opgewarmd zijn tot een temperatuur van ongeveer 150 °C. Het losse tonerpoeder wordt hierdoor als het ware gesmolten en hecht zich vast op het papier. Het opgewarmde papier gaat nu naar de opvangbak.

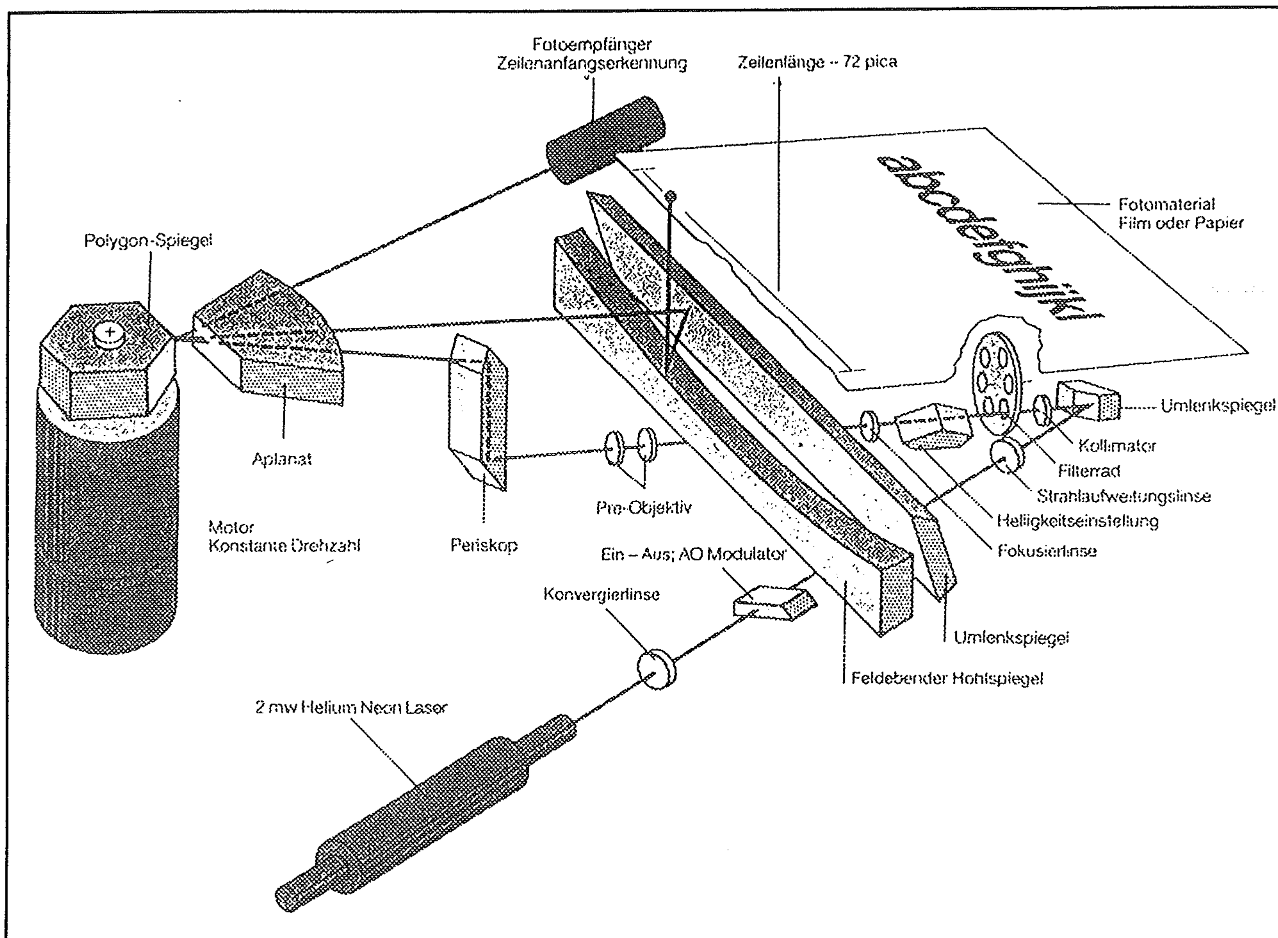
Het tonerpoeder

Het tonerpoeder bestaat uit een zeer fijne menging van 50 % ijzeroxyde, 35 % hars, 10 % verfpigmenten en 5 % additieven. Het ijzeroxyde is noodzakelijk om ervoor te zorgen dat het poeder goed aan de magnetische verdeelrol blijft kleven en gemakkelijk van deze rol naar de ongeladen delen van de drum springt. Het hars is verantwoordelijk voor het insmelten van de toner op het papier. De additieven bevatten onder andere zeer fijne deeltjes van een harde ceramische stof. Deze deeltjes worden gebruikt om het fotogevoelig oppervlak van de drum schoon te schuren, zodat papierdeeltjes die er blijven aan kleven binnen de kortste keren worden weg geschuurd.

Iedere fabrikant gebruikt een eigen mengverhouding, waarbij tegenwoordig bovendien nog eens een onderscheid wordt gemaakt tussen normale toner en zogenoemde "microtoner". Dit laatste poeder moet, althans volgens de fabrikanten, gebruikt worden bij laserprinters die een hoge resolutie hebben (800 dpi of meer). Het zou de kwaliteit van de afdruk aanzienlijk verbeteren. In de dagelijkse praktijk blijkt echter dat er nauwelijks een verschil is afdrukkwaliteit valt op te merken tussen gewone goedkope toner en de dure microtoner.

Dit naslagwerk wordt afgedrukt op een 1.000 dpi laserprinter, waarvoor de fabrikant zeer dure microtoner cassettes op de markt brengt. In plaats daarvan worden standaard LaserJet cartridges gebruikt en er is werkelijk niemand die enig verschil merkt!

19.11 Principes en werking van printers



Figuur 3/19.11-22: Het principe van de polygoon scan techniek.

Het belichtingssysteem

In figuur 3/19.11-21 werd als belichtingsbron een laser getekend. Nu zijn er echter in de loop der jaren vier belichtingssystemen ontwikkeld, waarvan er drie in feite helemaal niets te maken hebben met lasers. Desondanks worden ook printers die met deze systemen werken laserprinters genoemd! De vier systemen op een rijtje:

- de polygoon scan techniek;
- het LED-array;
- het LCD-array;
- de tonerjet-techniek.

De polygoon scan techniek

Bij deze techniek wordt met een laserbron gewerkt. Een laser is een lichtbron die een zeer dunne lichtstraal uitzendt. Bij "dun" moet niet aan millimeters worden gedacht, maar aan duizendsten van een millimeter. Het principe van deze belichtingstechniek is getekend in figuur 3/19.11-22.

Hart van het systeem is een helium/neonlaserbuis, met een vermogen van een paar mW. Deze zendt een continue zeer dunne (2/100 mm) lichtbundel uit, die eerst via een convergentielens nog meer wordt gebundeld. Nadien volgt de lichtmodulator. Hier wordt de elektronische informatie, afkomstig van de computer, omgezet in een pulserende lichtstraal. Via een heel ingewikkeld systeem van lenzen en spiegels belandt de met gegevens gemodu-

19.11 Principes en werking van printers

leerde lichtstraal uiteindelijk op de polygoon spiegel. Dit is het belangrijkste onderdeel van het optisch systeem. Zoals de naam reeds doet vermoeden, bestaat deze spiegel uit segmenten, in de meeste gevallen uit zes. De spiegel draait heel snel rond, toerentallen van 7.600 omwentelingen per minuut zijn geen uitzondering. De laserstraal valt in op de segmenten van deze spiegel en zal natuurlijk worden teruggekaatst, volgens de eenvoudige optische wet die zegt dat de invalshoek gelijk is aan de weerkaatsingshoek. Doordat de spiegel draait zal de weerkaatsingshoek echter lineair variëren. Het gevolg is dat de laserstraal een heen-en-weergaande beweging gaat uitvoeren. Op deze manier zal de straal lijn na lijn van de draaiende drum gaan scannen en de optische gegevens omzetten in een ladingsbeeld op de trommel. Het oppervlak van de trommel is in figuur 3/19.11-22 vlak voorgesteld, maar in realiteit is dit oppervlak natuurlijk gebogen. De door de polygoon spiegel teruggekaatste lichtstraal doorloopt nog eens een heleboel spiegels en lenzen alvorens terecht te komen op het oppervlak van de drum. Een belangrijk onderdeel van het systeem is de foto-ontvanger. Deze detecteert het begin van iedere lijnscan, zodat het systeem weet wanneer de belichtingseenheid begint met het scannen van een nieuwe lijn van de drum.

Hoge optische eisen

Aan de gehele optiek worden zeer hoge eisen gesteld. Dit geldt nog meer voor de polygoon spiegel. Deze bepaalt immers de nauwkeurigheid en gelijkmatigheid, waarmee de laserstraal het oppervlak van de drum belicht. Zelfs de geringste afwijking zorgt voor een duidelijk merkbare afwijking in de manier waarop het beeld op de drum en dus ook op het papier wordt geschreven. De afwijking in de vlakheid van deze spiegel moet kleiner zijn dan $^{6}/_{100.000}$ mm, de afwijking in hoeksnelheid kleiner dan 0,000.5 °.
Om een acceptabele printsnelheid te verkrijgen duurt het belichten van één punt op de trommel slechts ongeveer $^{1}/_{13.000.000}$ seconde.

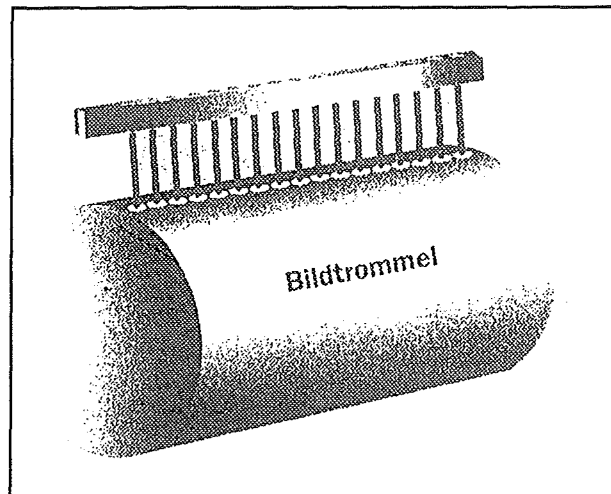
Nadeel van de polygoon scan techniek

De polygoon scan techniek heeft als belangrijkste nadeel dat het systeem heel ingewikkeld in elkaar zit en daardoor kostbaar is. Zolang men er echter niet in slaagde dunne laserbundels op een andere manier te produceren dan via een vrij groot uitgevallen helium/neon-laserbuis, was dit de enige manier om een laserprinter aan de praat te krijgen. Een ander nadeel van deze techniek is een effect dat bekend staat onder de naam "convergentie-fout". De polygoon spiegel staat opgesteld in het midden van de drum. De afstand die de laserstraal van de spiegel naar het oppervlak van de drum aflegt is echter niet constant. Als een punt op het midden van de drum wordt geschreven, zal de laserstraal een kleinere afstand afleggen dan wanneer een punt aan de linker of rechter uiteinde van de drum beschreven wordt. Nu is de laserstraal weliswaar heel dun en zal nauwelijks breder worden, maar toch is de lichtstip die wordt geproduceerd aan de uiteinden van de drum iets groter dan deze die geschreven wordt in het midden van de drum. Bij het printen van tekst heeft men van dit verschijnsel geen last. Als men echter een laserprinter gebruikt voor het afdrukken van gerasterde foto's, zal deze puntverdikking aan de uiteinden van de drum wel degelijk zichtbaar zijn op de afdruk.
Weliswaar kan men dit verschijnsel minimaliseren door een ingenieuze construc-

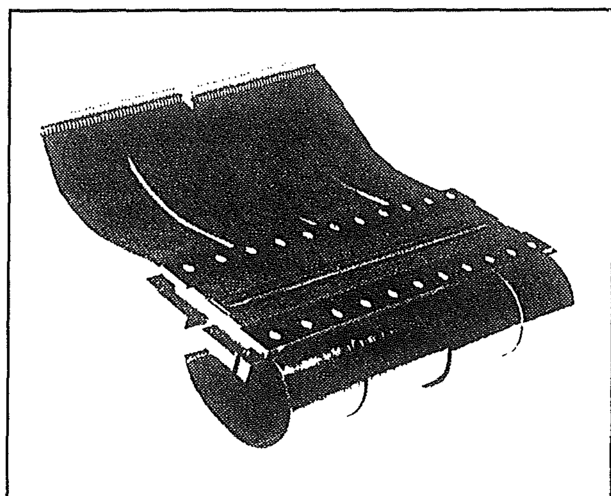
tie van lenzen en licht gebogen spiegels toe te passen (zie figuur 3/19.11-22). Maar de convergentie-fout blijft, hoe dan ook, bestaan!

Het LED-array

Tegenwoordig kan men een laserbron maken onder de vorm van een klein elektronisch onderdeel, een LASER-LED. Zo'n LASER-LED hoeft niet groter te zijn dan een/honderdste van een millimeter en het is niet zo'n probleem om duizenden van deze dioden keurig naast elkaar op een drager aan te brengen. Zo'n constructie noemt men een "array" en een dergelijk array, opgebouwd uit bijvoorbeeld 2.500 laserdioden, kan het gehele ingewikkelde optische systeem van de polygoon scan techniek vervangen!
In figuur 3/19.11-23 is het wel zeer eenvoudige principe van de LED-array belichtingstechniek geschetst. Een of meerdere LED-array's worden vlak boven de drum aangebracht. Gebruikt men meerdere array's, dan wordt er natuurlijk weer gebruik gemaakt van de reeds bekende dual-in-line staggered techniek, waarbij de individuele laser-dioden in ieder array iets van plaats verschoven zijn. De duizenden laser-LED's worden nu individueel aangestuurd, zodat in een keer een volledige beeldlijn op de drum wordt belicht. Op deze manier is men in staat uiterst compacte en goedkope laserprinters te fabriceren met een resolutie van 300 dpi.
Het zal duidelijk zijn dat de aansturingselektronica in een dergelijke printer vrij complex is. Maar elektronica is tegenwoordig zeer goedkoop te maken en de prijs van een dergelijk systeem valt in het niet als men het vergelijkt met de prijs van het zeer ingewikkelde mechanische systeem dat de polygoon-techniek nodig heeft.



Figuur 3/19.11-23: Het principe van een LED-array als belichtingseenheid.



Figuur 3/19.11-24: Een laser-LED array, zoals dit er in het echt uitziet.

Bovendien zijn LED-printers natuurlijk veel betrouwbaarder, omdat zij veel minder bewegende onderdelen bevatten.
In figuur 3/19.11-24 is de praktische uitvoering van een laser-LED array gefotografeerd. De eenheid meet 24 bij 6 cm^2 en wordt gemaakt met een ceramische drager als basis. Het array bestaat uit 20 monolytische elementen die ieder 128 LED's bevatten, gemaakt van gallium-arsenicum-fosfor als basismateriaal.

(wordt vervolgd)

19.11 Principes en werking van printers

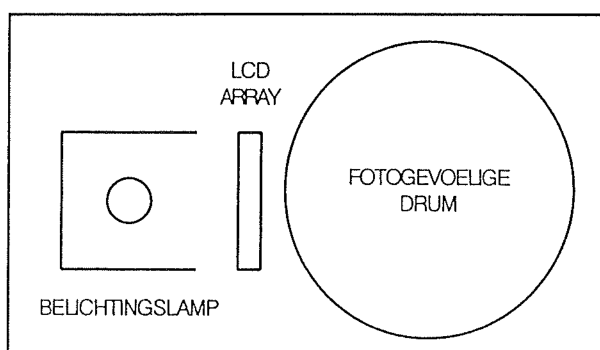
Ieder element wordt door een eigen chip aangestuurd, die niet alleen de stroombronnen bevat voor het aansturen van de LED's maar ook de schuifregisters en buffergeheugens om de gegevens op het juiste moment naar de LED's te sturen. De gegevens van een drumlijn worden over een 8 bit brede bus met een snelheid van 1 MB/s in de schuifregisters van de eenheid ingelezen. Nadien volgt een strobesignaal, waarmee de inhoud van alle schuifregisters in een keer naar de buffergeheugens wordt gecopieerd. Bij een volgende puls sturen de flip-flop's van deze geheugens de stroombronnen, zodat de LED's worden aangestuurd en gaan branden. De LED's zijn bij dit model, gefabriceerd door Telefunken, opgenomen in een raster van 85 μm, hetgeen een resolutie van iets minder dan 300 dpi oplevert.

Voordelen van de LASER-LED printers
Laserprinter die werken met LASER-LED array's hebben een aantal voordelen. De golflengte van het door de LED's uitgestraalde licht kan gemakkelijk aangepast worden aan de gevoeligheid van het fotogevoelig oppervlak van de drum.
Het gevolg is dat volstaan kan worden met kortere belichtingstijden, waardoor de snelheid van het printen stijgt.
Een tweede, reeds genoemd, voordeel is de eenvoud van de constructie en het totale gebrek aan mechanische onderdelen. Hierdoor kunnen dergelijke printers tegen zeer concurrerende prijzen op de markt worden gebracht, terwijl de bedrijfszekerheid op lange termijn uitstekend is. Laser-dioden hebben een hoog rendement, waardoor er weinig elektrische energie noodzakelijk is voor het belichten van de drum. Hierdoor werken dergelijke printers zuiniger dan de polygoon printers.
LED-printers hebben geen last van optische vervormingen, waardoor horizontaal over het papier lopende lijnen ook echt kaarsrecht worden geprint. Bij polygoon printers zullen rechte lijnen altijd een kleine, maar best merkbare tonvervorming hebben.
LED-printers kunnen gemakkelijk ontworpen worden voor het afdrukken op grote formaten. Er zijn reeds LED-printers ontwikkeld, die papierformaten tot A0 (84,1 x 118,9 cm^2) kunnen bedrukken. Dergelijke grote formaten kunnen, vanwege de beperkingen van het optische systeem, nooit met het polygoon-systeem bedrukt worden.

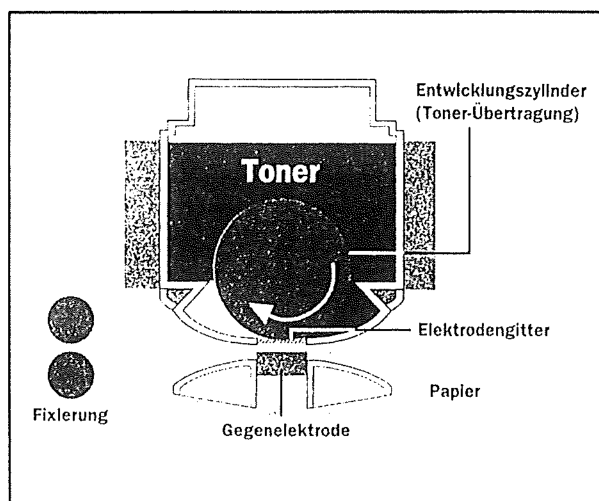
Het LCD-array
Voor dergelijke printers is de naam laserprinter helemaal misplaatst. Het LED-array wordt nu vervangen door een array dat is samengesteld uit duizenden LCD-celletjes. LCD is de afkorting van "Liquid Cristal Display" en deze technologie wordt bijvoorbeeld ook toegepast in polshorloges en de platte beeldschermen van draagbare computers. Een LCD-cel kan, onder de invloed van een elektrische spanning, ofwel invallend licht volledig blokkeren, ofwel het licht doorlaten. Het principe van deze belichtingstechniek is geschetst in figuur 3/19.11-25.
Het LCD-array wordt beschenen door een lamp, het licht van deze lamp wordt ofwel door de LCD-cellen gesperd, ofwel doorgelaten. In het laatstgenoemde geval valt het licht in op het fotogevoelig oppervlak van de drum.
De besturingselektronica voor een LCD-array is in grote lijnen vergelijkbaar met deze die gebruikt wordt bij een LED-array. Ook de voordelen van de LED-printer zijn van toepassing op de LCD-printer. Met één uitzondering, echter: het overschake-

19.11 Principes en werking van printers

len van licht blokkeren naar licht doorlaten gaat bij een LCD-cel vrij traag, zeker als men het vergelijkt met de snelheid waarmee een laser-diode in- en uitschakelt. De maximale printsnelheid van een LCD-printer is bijgevolg lager dan deze van een LED-printer.



Figuur 3/19.11-25: Het principe van printers die werken met een LCD-array.



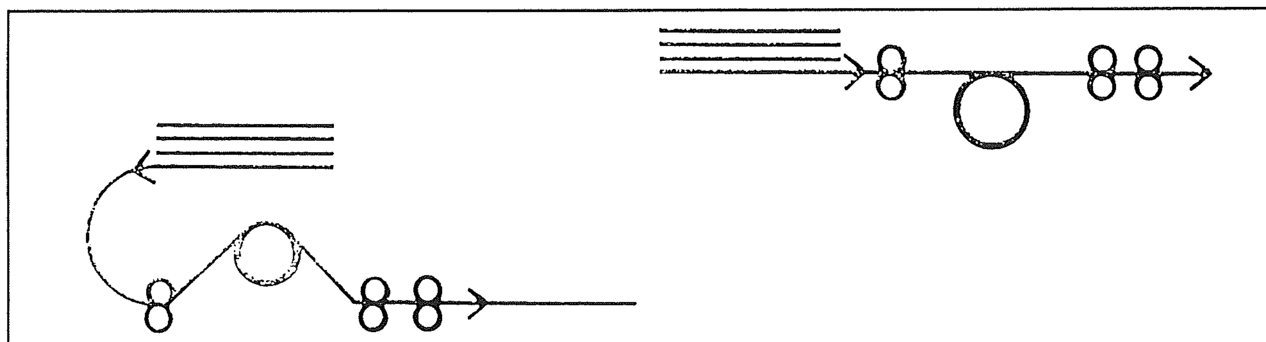
Figuur 3/19.11-26: Het werkingsprincipe van de tonerjet technologie.

De tonerjet techniek

Deze technologie is vrij nieuw en er zullen op dit moment weinig laserprinters op de markt zijn die ermee werken. Het systeem werd ontwikkeld door de Zweedse firma Array Printers AB en het productierecht werd verkocht aan het Duitse concern Telekom AG. Het revolutionaire van het systeem is dat de tussenfase van het opbouwen van een elektrostatisch beeld volledig wordt overgeslagen. Het papier wordt rechtstreeks vanuit het tonerreservoir bedrukt. Hoe dat kan volgt uit figuur 3/19.11-26.

Er is nog wel een soort drum aanwezig, maar die draait in het tonerreservoir. De enige taak van deze ontwikkelcilinder is een zeer dunne en zeer gelijkmatige laag toner over de gehele breedte van het te bedrukken papier te verkrijgen. Deze cilinder maakt contact met een zogenoemde roosterelektrode (Elektrodengitter). Deze elektrode bestaat uit een kunststoffolie, waarin duizenden uiterst kleine gaatjes zijn aangebracht. De wanden van al deze gaatjes zijn verguld. Dit gebeurt op dezelfde manier als waarmee de gaatjes in een printplaat worden doorgemetalliseerd. Deze gouden contactpunten worden aangesloten op de besturingselektronica. Het papier wordt onder de roosterelektrode doorgevoerd. Onder het papier is één massieve tegenelektrode aanwezig. Iedere gouden elektrode in ieder gaatje kan afzonderlijk onder spanning worden gezet. Daardoor kan er een spanningsverschil over het papier ontstaan. Dat spanningsverschil wekt een elektrostatisch veld op en dit veld trekt de moleculen van het tonerpoeder als het ware door de gaatjes naar het papier. Op deze manier ontstaat dus weer lijn voor lijn een beeld op het papier. Het op deze manier met tonerpoeder bedrukte papier gaat vervolgens weer door twee verwarmde fixeerrollen, waar het tonerpoeder op het papier gesmolten wordt.

19.11 Principes en werking van printers



Figuur 3/19.11-27: Het principieel verschil tussen gebogen en rechte papierdoorvoer.

Voordelen van de tonerjet technologie

Deze nieuwe printertechnologie wekt hoge verwachtingen. De roosterelektrode kan zoveel gaatjes bevatten, dat resoluties van 600 dpi zonder problemen mogelijk zijn. Bovendien is de dure fotogevoelige drum volstrekt overbodig. Het systeem werkt erg snel, zodat hoge papierdoorvoersnelheden verwacht kunnen worden. Helaas heeft deze technologie zich nog niet in de praktijk kunnen bewijzen. De allereerste "laser"-printer, die volgens deze technologie werkt en een resolutie heeft van 200 dpi, werd op de Duitse vakbeurs Cebit '95 aan het publiek voorgesteld.

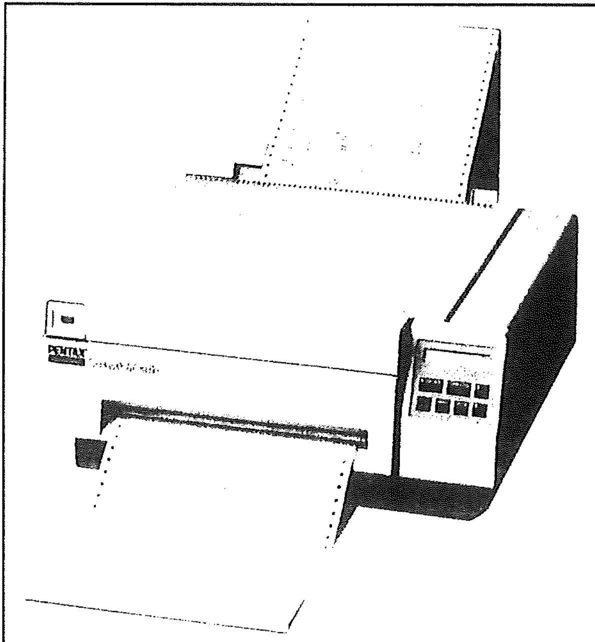
Papierdoorvoer

Bij laserprinters heeft men voor het transport van het papier door de printer twee systemen ontwikkeld. Deze zijn geschetst in figuur 3/19.11-27. Bij de gebogen papierdoorvoer maakt het papier tijdens zijn weg van voorraadbak naar opvangbak een bocht van 180 °. Het vel papier wordt, nadat het uit de voorraadbak is opgehaald, rond een rubber rol geleid, die het vel keert. Bij dergelijke printers zijn het dus de onderzijden van de vellen die worden bedrukt. Dat is van belang om te weten als men papier tweezijdig wil bedrukken. Het eenzijdig bedrukte papier moet dan met de bedrukte zijde naar boven in de voorraadbak worden gelegd!
Bij het tweede systeem, de rechte papierdoorvoer, verloopt het transport van het vel papier kaarsrecht. Het nadeel van dit systeem is dat de laserprinter veel dieper wordt, omdat voorraadbak, printmechanisme en opvangbak achter elkaar geplaatst moeten worden. Het grote voordeel is echter dat er geen mechanische spanningen op het papier worden uitgeoefend. Bovendien kan men zonder probleem dikker papier bedrukken, papier dat bij een gebogen papierdoorvoer onherroepelijk zou vastlopen.

Formulierenprinter

Laserprinters zijn typische losse vellen printers. Voor zakelijk/administratieve toepassingen heeft men echter grote behoefte aan apparatuur die kettingformulieren kan bedrukken. Van alle besproken typen is alleen de oude matrixtechnologie in staat om aan deze behoefte te voldoen. Gelukkig zijn er ook een aantal laserprinters verkrijgbaar, die in staat zijn kettingformulieren te verwerken. Een typisch voorbeeld van een dergelijk apparaat wordt voorgesteld in figuur 3/19.11-28. De Laserfold 240 van Pentax heeft een papiertransport, dat volledig is aangepast aan het verwerken van kettingformulie-

ren. De perforatiestroken passen in getande wieltjes en in feite gaat het inleggen van een nieuwe stapel papier net zo eenvoudig als bij een matrixprinter. Het enige probleem is dat de printer is ontworpen voor papier dat, van scheurrand tot scheurrand, precies 11 inch lang is, terwijl tegenwoordig meestal met kettingpapier van 12 inch wordt gewerkt. De printer, die alleen de standaard IBM karakterset kan afdrukken, doet dit razendsnel met een productie van 16 pagina's per minuut.



Figuur 3/19.11-28: De Pentax Laserfold 240, een laserprinter die kettingformulieren kan verwerken.

Klassen

Alle laserprinters zijn per definitie non-impact printers en werken pagina-georiënteerd.

Postscript?

Laserprinters zijn te koop in twee klassen. De eerste klasse, de goedkoopste, heeft geen eigen processor aan boord en moet de gegevens in printbare vorm aangeleverd krijgen door de in het systeem geïnstalleerde printer-driver.
De tweede soort is wat heet "Postscript-compatibel". Postscript (PS) is een pagina beschrijvingstaal (zie later) die de volledige inhoud van een pagina beschrijft onder de vorm van BASIC-achtige commando's. Deze commando's worden door een eigen processor in de printer geïnterpreteerd en omgezet in afdrukbare gegevens. Uiteraard zijn dergelijke laserprinters duurder, maar zij zijn universeel inzetbaar. Goedkope laserprinters kunnen geen Postscript-gegevens afdrukken, tenzij men een of andere softwarematige Postscript-emulator in het systeem opneemt, zoals het bekende "GoScript".

Goedkoop is vaak duurkoop

Dat geldt natuurlijk in zijn algemeenheid, maar zeker bij de aankoop van een Postscript-compatibele laserprinter.
Tegenwoordig worden er talloze PS-laserprinters aangeboden voor prijzen die lager zijn dan *f* 1.000,00. Op de kwaliteit van de afdruk is weinig aan te merken, zoals een demonstratie in de winkel aantoont. Men komt dan snel in de verleiding een dergelijke goedkope printer aan te schaffen. Toch zijn er een aantal haken en ogen, waarmee men maar beter rekening houdt! Dat heeft te maken met de elektronica in een PS-laserprinter, met name met de snelheid van de processor en de omvang van het geheugen.
PS-laserprinters bevatten vrij veel elektronica, waaronder een eigen processor en een eigen geheugen. De processor wordt video-controller of "RIP" genoemd, afkorting van "Raster Image Processor". Het woord "video" heeft niets te maken met de monitor van het computersysteem, maar alles met het feit dat die processor de inkomende Postscript-commando's ver-

19.11 Principes en werking van printers

taalt naar een patroon van elektronische “enen” en “nullen”, die nadien lijn na lijn op de drum worden belicht. Iedere elektronische “een” of “nul” komt overeen met een dotje, dat nadien al dan niet op de drum belicht zal worden. Deze techniek is vergelijkbaar met de manier waarop het beeld op een monitor lijn na lijn wordt opgebouwd en vandaar de naam video-processor. De RIP vertaalt de BASIC-achtige commando's van een hogere printertaal, zoals Postscript, op een identieke manier naar “enen” en “nullen”.
De processor stuurt de printgegevens niet rechtstreeks naar de drum, naar eerst naar het geheugen. In feite wordt dus in het geheugen een elektronisch schaduwbeeld opgebouwd van het beeld dat nadien door de drum op het papier wordt gezet. Die tussenstap is noodzakelijk, omdat de processor niet snel genoeg kan rekenen om de dot-gegevens in de door de drum vereiste snelheid aan te leveren. Bij het opslaan van de dot-gegevens in het geheugen worden datareductie technieken toegepast, zoals “RLE” oftewel “Run Lenght Encoding”. Een opeenvolging van 100 “nullen”, dus 100 niet te printen dot's, kan hierbij heel efficiënt worden gecodeerd met slechts twee bytes in het geheugen. Eerst nadat de volledige pagina op deze manier is omgezet in een patroon van “enen” en “nullen” start het eigenlijke printproces. De elektronica leest het geheugen lijn na lijn uit, decodeert de gecomprimeerde gegevens en stuurt de gegevens naar de laserstraalmodulator of het array. De snelheid van dit uitlezen is alleen afhankelijk van de snelheid waarmee de drum rond draait.

Het basisgeheugen

Het feit dat een PS-laserprinter maar eerst gaat printen als alle gegevens in het geheugen zitten, stelt bepaalde eisen aan de omvang van dit geheugen. Goedkope laserprinters worden uitgeleverd met een basisgeheugen van 512 kB. Nu is dit geheugen groot genoeg om een pagina tekst onder gecomprimeerde vorm te bevatten. Zo'n pagina bevat immers veel opeenvolgende dot's die niet geprint moeten worden en die kunnen flink gecomprimeerd worden. Anders wordt het echter als men ingewikkelde grafieken, tekeningen of gerasterde foto's wil afdrukken. Bij dergelijke gegevens komt het veel minder vaak voor dat een groot aantal opeenvolgende dot's niet geprint moet worden. De RLE-codering is dan niet zo efficiënt, met als gevolg dat het geheugen vol is alvorens de RIP er de hele pagina heeft ingelezen. In de meeste gevallen zal de printer dan weigeren een pagina te drukken, of verschijnt er slechts een halve pagina op het papier!

Het probleem van soft-fonts bij PS-printers

Een tweede probleem met de geheugenomvang doet zich voor als men zogenoemde “soft-fonts” bij een Postscript-compatibele laserprinter moet gebruiken. De RIP is standaard voorzien van een groot aantal PS-lettertypen en weet dus hoe hij deze moet omzetten naar een dot-patroon. Als men echter een lettertype gebruikt, dat de RIP niet kent, wordt dit lettertype standaard afgedrukt als Courier, de aloude schrijfmachineletter. Nu kan men natuurlijk de beschrijving van het lettertype in het geheugen van de printer laden, waardoor de RIP weet hoe hij deze letter op het papier vorm moet geven. Dat laden van een dergelijk soft-font kost echter geheugenruimte, ruimte die dan niet meer beschikbaar is voor het opbergen van de gegevens die geprint moeten worden.

19.11 Principes en werking van printers

Geheugen uitbouwen

Als men alleen teksten wil afdrukken in de standaard lettertypen zal het basisgeheugen van 512 kB groot genoeg zijn. Als men echter veel grafisch werk moet printen en bovendien gebruik wil maken van softfonts, doet men er heel onverstandig aan om een goedkope PS-laserprinter te kopen met een beperkt geheugen. Dan kan men beter wat meer geld uitgeven, zodat men een printer ter beschikking krijgt met een geheugen van 2 MB. Weliswaar leveren de meeste fabrikanten van laserprinters losse geheugenmodulen, die men in de printer kan inbouwen. Maar die modulen zijn in feite veel te duur, zodat besparen op geheugen eigenlijk duurkoop is.

De printsnelheid

De printsnelheid van een laserprinter is moeilijk te definiëren. Men moet hierbij namelijk rekening houden met twee verschillende verwerkingssnelheden:

- de snelheid waarmee de processor in staat is de binnenkomende gegevens te vertalen naar een rasterbeeld en dit beeld in het geheugen op te bergen;
- de snelheid waarmee deze gegevens de drum belichten en uiteindelijk het beeld op het papier vormen.

Als men dus beweert dat een laserprinter 15 pagina's per minuut kan afdrukken, zegt dit eigenlijk nog niets. Waarschijnlijk bedoelt men hiermee te zeggen dat deze printer in staat is van een bepaalde pagina, die in het geheugen is opgeslagen, 15 identieke afdrukken per minuut te maken. De afdruksnelheid wordt dan natuurlijk alleen maar bepaald door de snelheid van het optische systeem en de omwentelingssnelheid van de drum en beiden zijn best in staat een dergelijke hoge snelheid te halen.

Anders ligt de zaak als men 15 verschillende pagina's wil afdrukken. Dan wordt de printsnelheid voornamelijk bepaald door de snelheid waarmee de video-processor of de RIP in staat is de inkomende gegevens te verwerken en om te rekenen naar een rasterbeeld in het geheugen. Die snelheid ligt aanzienlijk lager dan de eerstgenoemde.

Goedkope laserprinters, met een trage processor, halen hier soms niet meer dan één pagina per vijf minuten!

Bij de aankoop van een zogenoemde "goedkope maar toch zeer snelle" laserprinter moet men dus goed informeren welke snelheid bedoeld wordt.

Voordelen van de tonertechnologie

Printers die werken met toner hebben een aantal voordelen op andere technologieën. Voor iedere afdruk wordt verse toner gebruikt, zodat de zwarting van iedere afdruk 100 % is. In vergelijking met thermo-transfer printers, die uiteraard hetzelfde voordeel hebben, is de kostprijs van een afdruk op een tonerprinter echter veel goedkoper. Er gaat immers vrij weinig van het poeder verloren en een tonercartridge heeft dus een zeer lange levensduur. In vergelijking met inktjet printers is de zwarting ook 100 % op goedkoop, vezelhoudend papier. De gesmolten toner dringt immers niet in het papier in, maar blijft als een laag op het papier kleven.

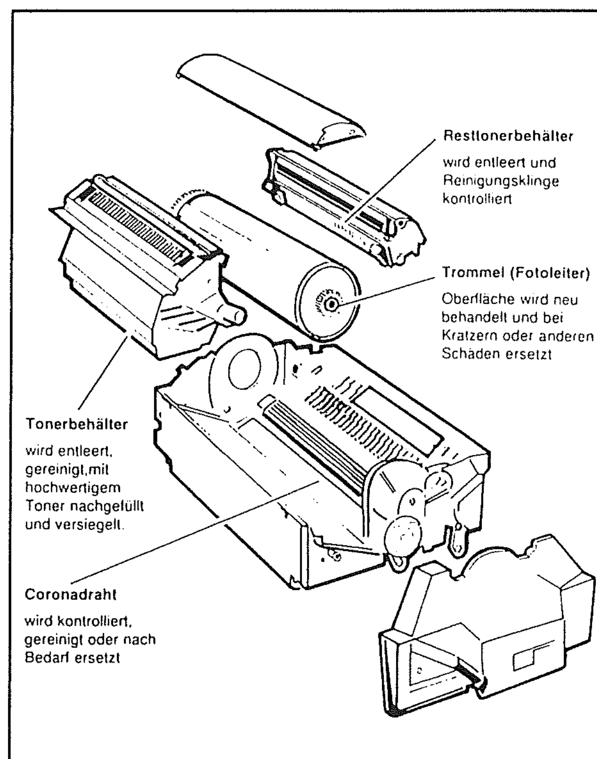
Problemen met uitvloeien van de inkt langs de vezelstructuur van het papier zijn niet aan de orde.

Een ander voordeel van deze technologie is, dat men tegenwoordig in staat is printers met zeer hoge resoluties voor een redelijke prijs op de markt te brengen. Waren enige jaren geleden 300 dpi printers de standaard op dit gebied, tegen-

19.11 Principes en werking van printers

woordig is het alles 600 dpi wat de klok slaat.
Laserprinters met een resolutie van 1.000 dpi zijn reeds betaalbaar in de handel!

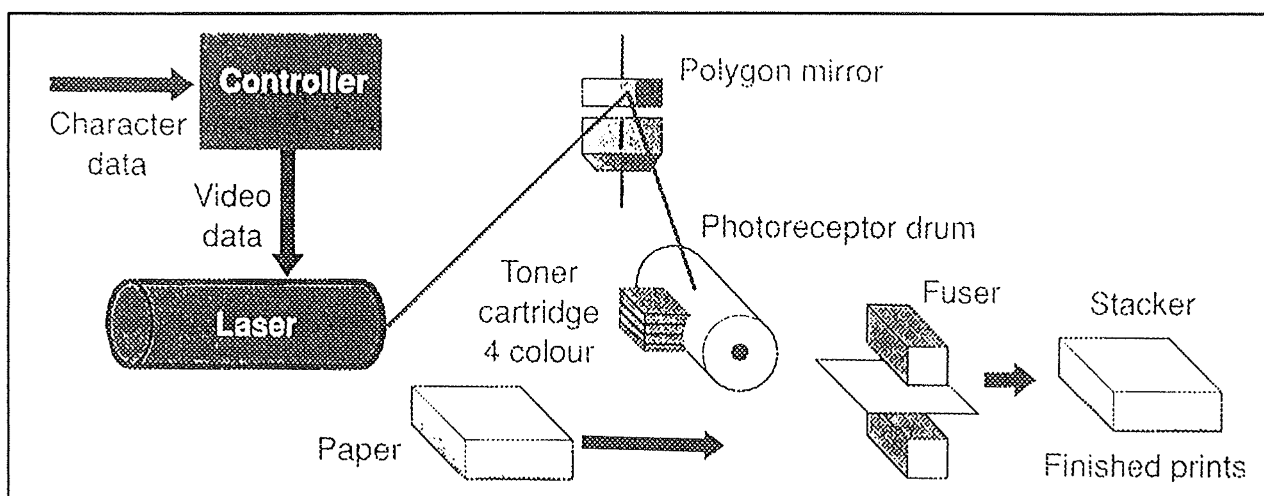
Nadelen van de tonertechnologie
Laserprinters die met een fotogevoelige drum werken hebben echter ook nadelen. Een van de belangrijkste nadelen van de "echte" laserprinters is dat de coronadraden, die de drum opladen, ook de lucht ioniseren. Het gevolg is dat dergelijke apparaten nogal wat ozon produceren, een gas met als chemische formule O_3. Deze "zuurstof in aangeslagen toestand", zoals ozon ook wel eens genoemd wordt, is in grote concentraties zeer slecht voor de gezondheid. In kantoorruimtes waar een aantal laserprinters aanstaan en bovendien ook nog eens kopieermachines, kan de ozonconcentratie tot onaanvaardbare hoogte stijgen. Gelukkig worden moderne laserprinters voorzien van een ozonfilter, dat tot taak heeft een groot deel van de uitgestoten ozon weer om te zetten in normale zuurstof O_2. De werking op lange termijn van dergelijke ozonfilters is echter zeer omstreden!
Een ander nadeel van de meeste laserprinters is dat men niet simpelweg de tonercartridge kan omwisselen als de printer zonder toner zit. Bij de meeste moderne laserprinters moet men de volledige print-unit vervangen, waar niet alleen nieuwe toner in zit, maar ook de primaire corona, de drum en het opvangmechanisme voor de toner die aan de drum blijft kleven. De samenstelling van een dergelijke unit is geschetst in figuur 3/19.11-29. Nu hebben moderne drum's een levensduur die veel langer is dan de gemiddelde vulling van een tonerreservoir. Men gooit dus veel bruikbaars weg!



Figuur 3/19.11-29: De samenstelling van de print-unit, die men bij de meeste laserprinters volledig moet vervangen als de toner op is.

Gelukkig bestaat er een levendige handel in tweedehandse cartridges. Lege print-units kan men gratis versturen naar een recyclingbedrijf, waar de units worden schoongemaakt en hervuld met nieuwe toner. De oude drum blijft dus behouden. Deze worden voor een zacht prijsje in de handel gebracht en voldoen net zo goed als nieuwe print-units.
Een ander nadeel van alle printers die werken met poedervormige toner is dat de uiterst kleine dotjes poeder die via de drum op het papier worden aangebracht, door het fixeren iets worden uitgesmeerd. Dit noemt men "puntverbreding" en dit is een eigenschap waar weinig tegen te beginnen valt. Dit valt niet erg op als men alleen teksten print.

19.11 Principes en werking van printers



Figuur 3/19.11-30: Het principe van een kleuren laserprinter met enkele drum.

Bij het afdrukken van rasters, bijvoorbeeld in foto's, blijkt duidelijk hoe die puntverbreding tot gevolg heeft dat de dotjes van donkere grijstinten volledig dichtlopen. Vandaar dat het geprinte beeld van een 1.800 dpi laserprinter toch niet te vergelijken is met dat van een digitale fotozetter met dezelfde resolutie. Bij een digitale fotozetter belicht de laserstraal immers rechtstreeks een fotogevoelige film, die nadien ontwikkeld en gefixeerd wordt. De uiterst fijne laserstraal belicht puntjes die bij ideale belichting net zo groot zijn als de middellijn van de laserstraal. Bij het belichten en fixeren verandert er niets aan de grootte van de dot's.

Kleuren laserprinters

Eén drum vier keer belichten

Laserprinters die in kleur kunnen printen zijn nog volop in ontwikkeling. Ondanks dat de leverbare apparatuur nog extreem duur is vallen de printresultaten best wel tegen. De "goedkopere" apparaten werken volgens het principe dat in figuur 3/19.11-30 geschetst is. Naast de fotogevoelige drum staan nu vier tonerreservoirs, gevuld met respectievelijk cyaan, magenta, gele en zwarte toner. De ingebouwde controller ontleedt de binnenkomende gegevens naar hun kleursamenstelling en bouwt vier elektronische beelden op, een voor ieder van de aanwezige tonerkleuren. Al deze gegevens worden opgeslagen in het geheugen, dat dus uiteraard veel groter moet zijn dan bij een monochrome printer.

Geheugens van 16 MB zijn eerder regel dan uitzondering!

Nadat de elektronische voorbereidingen klaar zijn, wordt de drum in vier fasen belicht. Bij de eerste rotatie schrijft de laser bijvoorbeeld het cyaan beeld op de drum, waarna het cyaankleurige tonerpoeder op de rol wordt aangebracht. Bij de drie volgende rotaties van de drum worden de drie overige kleuren belicht en neemt de drum het tonerpoeder van deze kleuren op. Bij de vijfde omwenteling van de drum wordt het papier aangevoerd en wordt het volledige kleurenbeeld in één gang overgedragen op het papier. Nadien volgt de fixering door middel van de twee warme rollen.

Dure drum

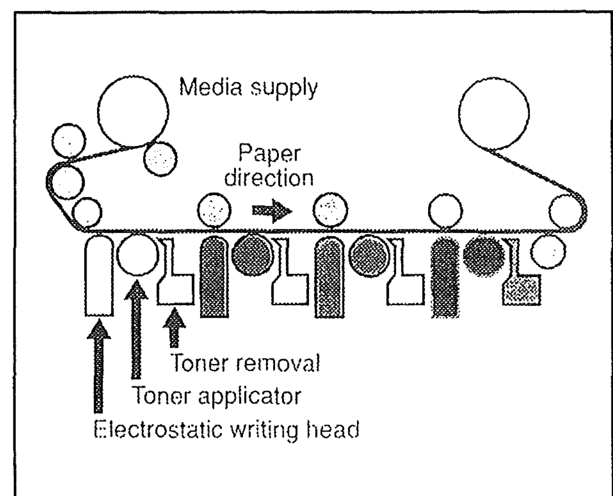
Een van de belangrijkste kostenposten bij een kleuren laserprinter is de drum. Bij monochrome laserprinters kan men werken met een drum met een omtrek van slechts 7,5 centimeter. Het volledige beeld wordt dan in vier opeenvolgende omwentelingen op de drum belicht en op het papier overgedragen. Bij een kleuren laserprinter is dit, vanwege het toegepaste principe, onmogelijk. Het volledige beeld moet eerst op het oppervlak van de drum geschreven worden. Vandaar dat de omtrek van de drum bij een A4-printer minstens gelijk moet zijn aan 30 cm.

Nadelen van de enkele drum technologie

Kleuren laserprinters die met één drum werken, hebben een aantal nadelen. Problemen zijn de perfecte passing van de deelbeelden, het zogenoemde "register" en de vervuiling van de kleuren bij het aanbrengen van de vier deelbeelden op de drum. Deze problemen zijn nog niet bevredigend opgelost.

Vier-drum systemen

Naast de kleurenlasers met één drum zijn er ook systemen ontwikkeld die werken volgens het principe dat ook bij kleuren offset drukpersen wordt gehanteerd. Zoals uit figuur 3/19.11-31 blijkt, zijn er nu vier afzonderlijke drukwerken aanwezig, ieder met eigen drum en belichtingseenheid. Het zal duidelijk zijn dat dergelijke apparaten extreem prijzig zijn en alleen rendabel zijn bij het verwerken van zeer hoge oplages. Toch worden dergelijke "elektronische drukpersen" een grote toekomst voorspeld en men verwacht dat zij binnen vijf jaar een ernstig te nemen concurrent zullen vormen voor de traditionele klein formaat offset persen. Het grote voordeel van dergelijke kleurenmachnines is dat men heel snel van de ene naar de volgende opdracht kan omschakelen, zonder het tijdrovende gedoe van het wisselen van offsetplaten en schoonmaken van de rubber doeken. Bovendien lenen zich dergelijke systemen voor "printing on demand" (POD) en "gepersonaliseerd drukwerk", waarbij de oplages tot op het vel worden afgestemd op de momentele behoeften van de opdrachtgever.



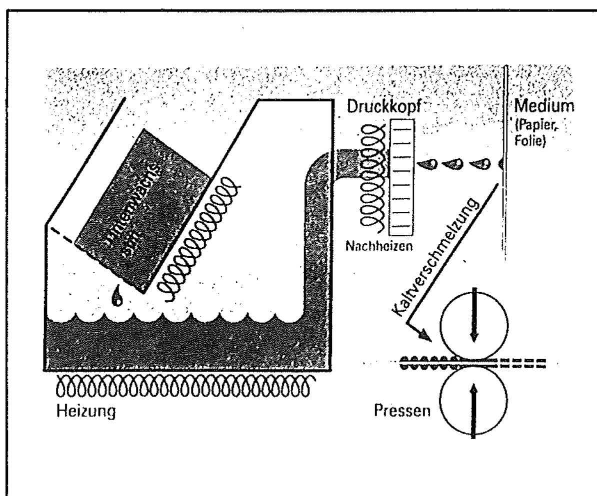
Figuur 3/19.11-31: Het principe van een kleuren laserprinter met vier drukwerken.

Kleuren thermal wax printers

Het principe

Thermal wax printers zijn gebaseerd op het principe van de inktjet printers. Groot probleem bij het inktjet procédé is dat de vloeibare inkt in het papier trekt, waardoor speciaal papier gebruikt moet worden om briljante kleurenafdrukken te krijgen. Thermal wax printers hebben dit nadeel niet en kunnen ook op normaal

papier schitterende kleurenafbeeldingen produceren. Er wordt nu niet gewerkt met vloeibare inkt, maar met ***transparante*** kleurstoffen die zeer fijn zijn gemengd met een soort was. Zoals uit figuur 3/19.11-32 volgt, worden de wasstaafjes (ook nu worden vier staafjes gebruikt in de standaard kleuren) eerst gesmolten in een reservoir. De vloeibare was wordt naar de printkop geleid, daar nog eens extra verwarmd tot de was dun vloeibaar is en via kleine gaatjes op het papier gespoten. Dat papier is koud, met als gevolg dat de uiterst kleine druppeltjes vloeibare was onmiddellijk stollen als zij in contact komen met het papier. De gestolde druppeltjes was bijven dus op het papier liggen, zodat er geen kans is dat de kleurstof zich via de vezelstructuur van het papier gaat verspreiden. Het bedrukte vel papier wordt nadien gefixeerd. Dat gebeurt nu niet met verwarmde rollen zoals bij laserprinters, want dan zou de vloeibare was toch weer in het papier kunnen dringen, maar door druk.



Figuur 3/19.11-32: Het principe van de thermal wax printers: vaste wasstaafjes worden in twee fasen gesmolten en dun vloeibaar op het papier gespoten.

Het bedrukte papier wordt tussen twee metalen rollen doorgevoerd, die met grote kracht tegen elkaar gedrukt worden. De druk die op het papier wordt uitgeoefend zorgt er nu voor dat de kleine gestolde wasdruppeltjes platgedrukt worden. Het grote voordeel hiervan is dat de typische rasterstructuur, die men bij alle tot nu toe besproken printertechnologieën aantreft, wegvalt. De afzonderlijke dotjes zijn niet meer als dusdanig te herkennen, waardoor het idee ontstaat alsof de afdruk ongerasterd is. Bovendien zullen door het koudwalsen de wasdruppeltjes van verschillende kleuren iets door elkaar gemengd worden. De gebruikte kleurstoffen zijn transparant, waardoor zeer mooie mengkleuren ontstaan.
Het is voornamelijk de firma Tektronix, die zich gespecialiseerd heeft in de thermal wax technologie. Er wordt een reeks zogenoemde "Phaser"-printers op de markt gebracht, die volgens deze technologie werken en uitstekend kleurendrukwerk kunnen produceren. Helaas zijn de prijzen van dergelijke apparaten nog steeds te hoog voor algemeen gebruik op het kantoor en in de kleine studio. Opgemerkt moet echter worden dat sommige typen uit de "Phaser"-serie niet met thermische was werken, maar volgens de sublimatie technologie of de thermo-transfer techniek.

Kleurstof sublimatie printers

Sublimeren

De meeste stoffen hebben drie zogenoemde agregatietoestanden: vast, vloeibaar en gasvormig. Door het verhitten van

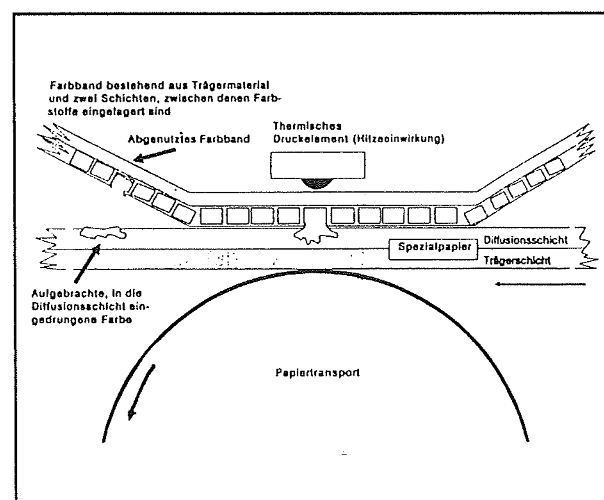
19.11 Principes en werking van printers

een vaste stof wordt deze eerst vloeibaar (het smeltpunt) en gaat nadien over in gasvormige toestand (het damppunt). Het smeltpunt van bijvoorbeeld water ligt op 0 °C, het damppunt op 100 °C. Nu zijn er bepaalde stoffen waarvan het smelt- en damppunt heel dicht bij elkaar liggen. Door een dergelijke stof in vaste toestand plotseling te verwarmen, zal de stof onmiddellijk verdampen, waarbij de vloeibare fase wordt overgeslagen. Een dergelijk proces, waarbij een stof opeens van vaste toestand in gasvormige toestand overgaat, noemt men sublimatie.

Sublimatie printers

Sublimatie printers maken gebruik van het beschreven fysische verschijnsel. Zoals uit figuur 3/19.11-33 volgt, lijkt het principe erg op dat van de thermo-transfer printers. Ook nu is er een rol printerfolie aanwezig, ingedeeld in opeenvolgende paginagrote vellen in de vier basiskleuren. Ook nu staat er achter deze rol een array met verwarmingselementen. Maar daarmee houden de overeenkomsten op! De inkt zit in vaste vorm opgesloten in kleine cellen in de rol printerfolie. Deze cellen worden plotseling verhit, waardoor de kleurstof opeens verdampt. Door de grote uitzetting die hiervan het gevolg is, explodeert de cel als het ware en spuit een klein wolkje ***gasvormige*** kleurstof naar het papier. Het papier staat niet in mechanisch contact met de printerfolie, er is dus een klein spleetje tussen folie en papier aanwezig. Het papier heeft een speciale samenstelling, het bestaat namelijk uit een drager met daarop een speciale diffusielaag. Deze laag is poreus, zodat zij in staat is de gasvormige moleculen van de weggeschoten kleurstof in zich op te nemen.

Het resultaat van een en ander is dat er geen scherp begrensde inktdot op het papier verschijnt, maar eerder een onregelmatig gevormd vlekje, met wazige randen. Vanwege dit feit is een sublematie printer eigenlijk helemaal niet geschikt voor het afdrukken van bijvoorbeeld zwarte teksten. De randen van de letters zijn veel te onscherp! Het principe is echter uitermate geschikt voor het afdrukken van kleurenfoto's. Ook nu wordt het papier vier keer door de printer gevoerd, waarbij telkens een andere kleurstrook van de printerfolie wordt gebruikt. Doordat de gasvormige wolkjes kleurstof in de poreuze bovenlaag van het papier indringen, ontstaan mooie mengkleuren. Bovendien zorgen de wazige randen ervoor dat de naast elkaar geprinte kleurdotjes netjes in elkaar overvloeien, zodat echt de indruk ontstaat van een kleurenfoto, zonder dat een hinderlijk waarneembaar dotraster aanwezig is.



Figuur 3/19.11-33: Het principe van de sublematie printer.

Voor- en nadelen

Sublimatie printers zijn op dit moment de enige kleurenprinters die in staat zijn foto's op een dergelijke manier af te drukken, dat het verschil met een "natte" kleu-

renfoto, gemaakt van een negatief, niet waarneembaar is. Maar, zoals gezegd, voor het afdrukken van lijntekeningen en teksten zijn deze printers niet zo ideaal. Een ander nadeel is de hoge prijs, zowel van het apparaat zélf als van de verbruikte materialen. Het afdrukken van een A4 pagina kost ongeveer tien gulden. Deze printers hebben dus de weg naar het gemiddelde kleine kantoor nog niet gevonden, maar zijn ideaal voor grafische studio's die kleurechte proefprinten van ontwerpen aan de klanten willen tonen.
Een ander voordeel is dat de afdrukken, gemaakt op een sublematie printer, vrij krasvast zijn.
Dit is natuurlijk een rechtstreeks gevolg van het feit dat de kleurstof in gasvormige toestand binnendringt in de diffusielaag van het papier en dus niet op het papier blijft liggen. Dit voordeel geldt zeer zeker als men de krasvastheid van een sublematie-afdruk vergelijkt met deze van een thermal-transfer afdruk.
Een nadeel is dat sublimatie-afdrukken niet bestand zijn tegen ultraviolette belichting. Afdrukken, gemaakt met dergelijke printers, verbleken zonder speciale maatregelen al na een paar maanden. Afdrukken die voor het nageslacht bewaard moeten blijven, moeten dus lichtdicht worden opgeborgen, behandeld met speciale spray's die UV blokkeren of worden opgehangen onder een raampje UV-3 plexiglas.

Opmerking
Kleurstof sublimatie printers worden soms ook aangeprezen als "DS/T2 printers". Deze afkorting slaat op het Engelse "Dye Sublimation, Thermal Transfer", een omschrijving die het werkingsprincipe van dergelijke printers kernachtig samenvat.

Klassen
Het zal wel zonder nadere toelichting duidelijk zijn dat kleurstof sublimatie printers behoren tot de non-impact printers en pagina-georiënteerd zijn.

Ionen printers

Inleiding
Het ionen-principe is vrij nieuwe en er zijn nog maar weinig printers beschikbaar die volgens dit principe werken. Het procédé werd in 1991 door de Japanse Nippon Telegraph and Telephone Corporation (NTT) ontwikkeld en een printer met een resolutie van 200 dpi werd reeds aan de pers voorgesteld. Het revolutionaire aan het procédé is dat men amplitudemodulatie kan toepassen. Alle tot nu toe behandelde printersystemen zijn alleen in staat of een dotje met een bepaalde diameter op het papier te zetten of dit niet te doen. Voor het afdrukken van teksten en lijntekeningen werkt dit uitstekend, maar die constante dot technologie is niet zo geschikt voor het afdrukken van foto's (zie later). Bij amplitudemodulatie kan men de grootte van de af te drukken dot op een heel eenvoudige manier variëren. Dat geeft enorm veel mogelijkheden in de rastertechnologie en bij het samenstellen van kleuren.

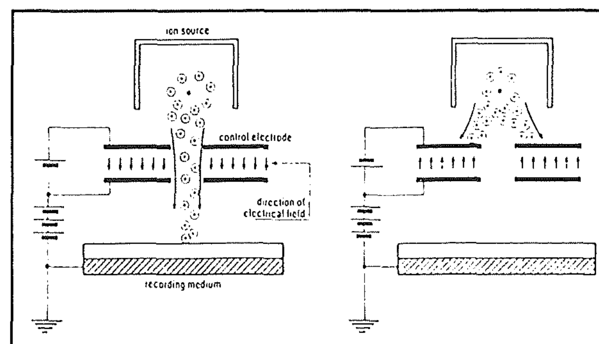
Amplitudemodulatie
Bij het prototype van de printer kan men de grootte van de dot in 256 stappen variëren. Dat betekent dat men zonder moeilijke softwarematige procedures (alweer, zie later) 256 verschillende grijstinten op het papier kan afdrukken. Het gevolg is dat men schitterende halftoon foto's kan printen. Past men de technolo-

19.11 Principes en werking van printers

gie toe op kleur, dan kan men de intensiteit van ieder van de drie basiskleuren in 256 stappen instellen. Dat betekent dat men in totaal 256 tinten cyaan maal 256 tinten magenta maal 256 tinten geel met elkaar kan mengen tot in totaal 16,7 miljoen mengkleuren op het papier! Dat wat op ieder standaard SVGA-beeldscherm reeds jaren mogelijk is, wordt dus voor het eerst ook mogelijk op papier! De professionele 24 bit kleurstandaard van DTP-omgevingen is hiermee ook bij printers een feit geworden.

Ionen

Een atoom is opgebouwd uit een elektrisch positief geladen kern, waar een aantal elektronen omheen cirkelen die negatief geladen zijn. De natuur heeft het zo geregeld dat een atoom elektrisch neutraal is. De positieve lading van de kern wordt gecompenseerd door de even grote negatieve lading van de elektronen. Nu is het met eenvoudige technieken mogelijk een of meerdere elektronen uit een atoom te verwijderen. Het achterblijvende, geamputeerde atoom heeft dan een overschot aan positieve lading en wordt een ion genoemd. Elektrisch geladen stoffen kunnen echter aangetrokken of afgestoten worden door een extern elektrisch veld. Ook die positieve ionen kunnen dus aangetrokken of afgestoten worden. Aantrekken of afstoten veroorzaakt een beweging.
Men kan dus die positieve ionen van de ene plaats (bijvoorbeeld een kleurstof reservoir) naar een andere plaats (bijvoorbeeld een vel papier) transporteren. Als men er nu voor zorgt dat het vel papier negatief opgeladen is, zullen de positieve ionen weer hun geamputeerde negatieve lading uit het papier opnemen en worden weer neutrale atomen. Maar ondertussen heeft men wél een minuscuul druppeltje kleurstof op het papier aangebracht!



Figuur 3/19.11-34: Het principe van ionen printers.

Het werkingsprincipe

Het werkingsprincipe van ionen printers is weergegeven in figuur 3/19.11-34. Hart van het systeem is een ionenbron, een metalen behuizing waarin een centrale elektrode is opgenomen. Door tussen de behuizing en de centrale elektrode een grote gelijkspanning te zetten wordt alles dat in het vat aanwezig is geïoniseerd. In de behuizing is de kleurstof aanwezig, die dus door het grote elektrisch veld wordt omgezet in positieve ionen. Onder de behuizing zijn twee elektroden aangebracht, voorzien van een gaatje met een middellijn van 10 µm. Tussen beide elektroden wordt een gelijkspanning gezet, die omgepoold kan worden. Dat wil zeggen dat in de ene toestand de onderste elektrode positief is ten opzichte van de bovenste en in de andere toestand de onderste elektrode negatief is en opzichte van de bovenste. Onder de elektroden zit het papier, dat op een grote negatieve spanning wordt gezet ten opzichte van het elektrodensysteem.
In de linker tekening is weergegeven wat er gebeurt als de bovenste elektrode positief wordt ingesteld. De positieve inkt-

ionen worden aangetrokken, gaan door het gaatje en komen op het negatief geladen papier terecht. Hier nemen de inkt-ionen negatieve lading op en transformeren zich tot neutrale atomen. In de rechter tekening is de situatie geschetst die ontstaat als de spanning op de twee elektroden wordt omgepoold. De positieve ionen worden nu aangetrokken door de negatieve lading van de bovenste elektrode en gaan dus niet via het gaatje naar het papier. Op deze eenvoudige manier kan men dus een dotje inkt op het papier printen.
Er is echter meer. De hoeveelheid positieve inkt-ionen die via het gaatje naar het papier vliegen is in te stellen door de grootte van de spanning tussen beide elektroden. Hoe groter deze spanning, hoe meer ionen worden aangetrokken en hoe meer inkt-atomen uiteindelijk op het papier terecht komen. Op deze manier kan men dus de hoeveelheid inkt, die plaatselijk op het papier terecht komt, moduleren. Hoe meer ionen door de elektroden naar het papier ontsnappen, hoe groter het dotje inkt zal zijn.

Technologische problemen
Het principe van de ionen printers is veelbelovend, maar toch moeten er nog heel wat problemen opgelost worden alvorens deze apparaten in ieder kantoor zullen staan. Een van de grootste problemen is dat ionen heel mooi gestuurd kunnen worden door elektrische velden in het absolute luchtledige. Nu is dat luchtledige bij een printer wel niet te bereiken. Stuurt men ionen door lucht, dan botsen de ionen tegen de luchtmoleculen, waardoor zij een zigzaggende beweging gaan uitvoeren. Het gevolg is dat de op het papier geprinte inktdot toch vrij groot is, dit ondanks het feit dat de ionen door een gaatje met een diameter van slechts een/honderdste van een millimeter ontsnappen. Vandaar dat de maximale resolutie die op dit moment bereikbaar is, slechts 200 dpi bedraagt. Men verwacht echter dat men dit probleem kan oplossen, waardoor de tegenwoordig als normaal ervaren resoluties van 300 tot 600 dpi bereikbaar zijn.
Als men hiertoe in staat is, belooft de ionen technologie een zeer zware concurrent te worden van alle bestaande printer technologieën.
Met name de mogelijkheid om modulatie van de inktdot toe te passen belooft prachtige resultaten bij het afdrukken van zwart/wit- en kleurenfoto's.

Nat-elektrostatische printers

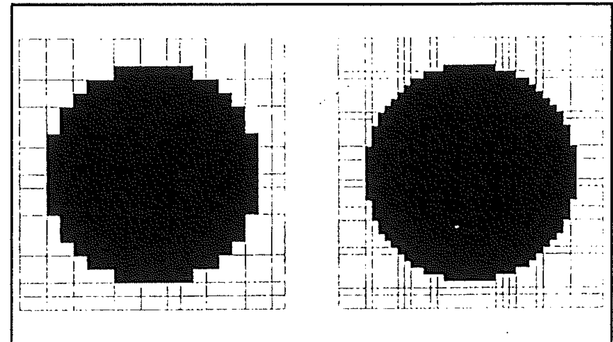
Professionele printers
Tot slot een printer technologie, die wel nooit gemeengoed zal worden in het kantoor. Bij de nat-elektrostatische printers wordt gewerkt met vloeibare toners. De kleurstoffen zijn verwerkt in een emulsievloeistof. In een emulsie wordt gebruik gemaakt van een vloeibare drager, waarin de actieve stoffen onder de vorm van zeer kleine druppeltjes als het ware zweven. De bekendste emulsie is zonder enige twijfel mayonaise. Het medium, waarop gedrukt moet worden, wordt elektrostatisch opgeladen, bijvoorbeeld door een laserstraal. Nadien wordt het medium door één bad emulsie (zwartdruk) of door vier baden (kleurendruk) gevoerd. Overal waar op het medium elektrische lading aanwezig is, worden de kleine druppeltjes uit de emulsie naar het papier getrokken. Na-

dien wordt het papier gedroogd, waarbij meteen de tonerdruppeltjes op het papier worden gefixeerd. Het voordeel van deze technologie is dat men in staat is printers te bouwen die zeer grote papierformaten kunnen bedrukken.
De haalbare resolutie bedraagt 400 dpi. Nadelen zijn dat vrij veel giftige dampen vrijkomen en dat men de nodige aandacht moet besteden aan het afzuigen van deze dampen. Vandaar dat dit soort printers alleen in gespecialiseerde drukkerijen worden toegepast, waarbij met hoge snelheid grote oplagen van elektronische gegevens moeten worden geprint.

Printer enhancement technology

Inleiding

Op de ionen printers na, leveren alle besproken technologieën in principe dot's af met een constante grootte. De grootte van deze dot's wordt bepaald door de diameter van een printnaald, van een laserstraal of door de grootte van een gaatje in een printkop.
Nu heeft die constante dotgrootte als nadeel dat de randen van hetgeen geprint wordt trapvormig verlopen. Een cirkel kan dus niet echt 100 % rond geprint worden, maar zal altijd bestaan uit een trapvormige benadering van de ideale omtrek. Dit wordt fraai toegelicht in figuur 3/19.11-35. Bij 300 dpi is dit verschijnsel zelfs duidelijk met het blote oog op te merken, maar ook bij 600 dpi valt nog steeds op dat het printbeeld iets niet heeft dat een afdruk van een dure hogeresolutie digitale fotozetter wel heeft: trapvrije kanten.



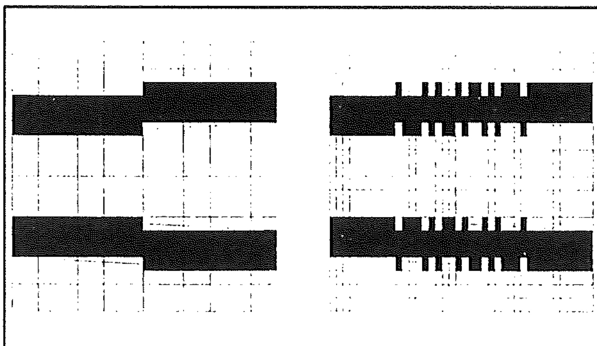
Figuur 3/19.11-35: Een ideale cirkel, geprint met 300 dpi (links) en met 600 dpi (rechts).

Diverse fabrikanten hebben methoden ontwikkeld om dit verschijnsel te minimaliseren. Deze technieken kunnen worden samengevat onder het algemene begrip "printer enhancement technology" oftewel "PET". Door het toepassen van deze technieken kan de resolutie van een laserprinter schijnbaar verhoogd worden. Let op het woord schijnbaar: er wordt nog steeds gebruik gemaakt van dezelfde printmechanismen met een mechanische resolutie van 300 of 600 dpi. Door dat printmechanisme echter softwarematig op een heel ingenieuze manier aan te sturen lijkt het erop alsof het geprinte beeld een veel hogere resolutie heeft dan het mechanisme.

TurboRes van LaserMaster

Door het Amerikaanse bedrijf LaserMaster werd een PET-systeem ontwikkeld, dat "TurboRes" wordt genoemd. De met dit systeem uitgeruste laserprinters hebben een mechanische resolutie van 400 dpi, maar worden door deze technologie opgewaardeerd tot een schijnbare 1.000 dpi. Het principe van TurboRes is vrij ingewikkeld en wordt aan de hand van figuur 3/19.11-36 toegelicht. In deze extreem uitvergrote tekening is links voorgesteld hoe licht hellende lijnen worden afge-

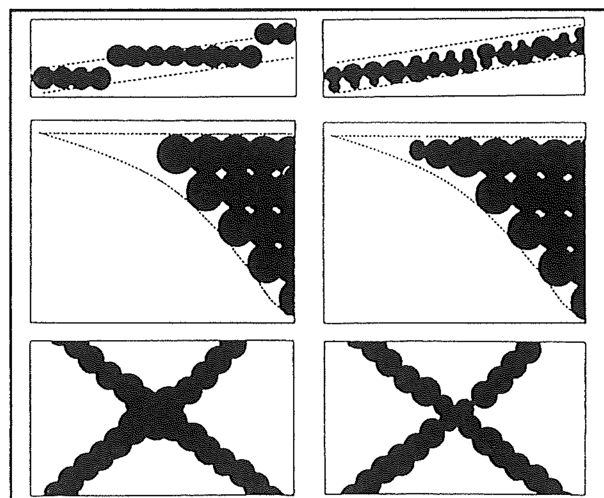
drukt op normale 400 dpi mechanismen. Men herkent duidelijk één van de trapjes, waarmee de licht hellende lijnen worden geprint. Het TurboRes-systeem heeft software aan boord die een dergelijke licht hellende lijn herkent. In plaats van nu alleen een trapvormige benadering van de lijn op papier te printen worden in de omgeving van een trapovergang een aantal dotjes extra geprint (rechter afbeelding). In deze extreme uitvergroting lijkt het alsof het TurboRes-resultaat echt niet beter is dan het normale resultaat, maar zelfs veel rafeliger op het papier verschijnt. Op ware grootte bekeken, echter, zorgen de eenzame dotjes er voor dat de trapvormige overgangen in de lijn veel minder opvallen en het lijkt dus alsof de lijn rechter verloopt.



Figuur 3/19.11-36: Het TurboRes-principe van LaserMaster, toegelicht aan de hand van het printen van licht hellende lijnen.

RET van Hewlett Pakard

Ook HP heeft in zijn serie Laserjet-III laserprinters een PET-systeem toegepast. HP noemt dit "RET", de afkorting van "Resolution Enhancement Technology". RET bestaat uit een extra stuk software dat in het bedrijfssysteem van de printer wordt ondergebracht en er voor zorgt dat de diameter van de dot's variabel wordt. De software moduleert de helderheid van de laserstraal. Hoe helderder deze straal hoe meer lading er ter plekke op de drum wordt geneutraliseerd en hoe meer tonerpoeder dit punt zal aantrekken. Hoe zwakker de intensiteit van de laserstraal, hoe minder lading er op de drum geneutraliseerd wordt en hoe minder tonerpoeder er aan de drum blijft kleven. De intensiteit van de laserstraal kan in vijf stappen ingesteld worden, namelijk op 20, 40, 60, 80 en 100 % van de maximale intensiteit. Iedere dot in het 300 dpi raster kan dus vijf verschillende diameters krijgen. Wat dit voor gevolg heeft op het printresultaat blijkt overtuigend uit de drie voorbeelden in figuur 3/19.11-37. Dotjes die normaal niet geprint zouden worden, omdat zij buiten de begrenzing van de te printen omtrek vallen, kunnen nu wel geprint worden, evenwel met een kleinere diameter. Hierdoor wordt de vloeiende omtrek veel beter benaderd en valt de trapvormige structuur veel minder op. Let er echter wel op dat ook nu de resolutie niet verhoogd wordt. Er worden wel meer dotjes geprint, maar deze zitten nog steeds in een raster van 300 dpi!



Figuur 3/19.11-37: Drie voorbeelden waaruit het effect van RET blijkt: links zonder RET, rechts met RET.

19.11 Principes en werking van printers

Blijft natuurlijk de vraag hoe de software in staat is te bepalen welke dot's kleiner afgedrukt moeten worden en welke niet. Hiervoor heeft Hewlett Packard een speciale ASIC ontwikkeld, een zeer ingewikkelde logische schakeling. Deze interpreteert de beeldinformatie die in het geheugen van de printer aanwezig is. De ASIC onderzoekt telkens een blokje dot's van 2.560 beeldpunten in de breedte en 6 beeldpunten in de hoogte. Vervolgens worden van ieder beeldpunt in zo'n blok de 49 omliggende punten geanalyseerd. Via ingewikkelde softwarematig gestuurde patroonherkennings-algoritmen worden een aantal standaard patronen, die blijkbaar vaak voorkomen bij het printen van tekst en lijnen, herkend. Is een bepaald patroon herkend, dat weet de software hoe dit patroon beter ingevuld kan worden door het extra printen van kleiner-dan-normale dot's. Het is vervolgens niet zo moeilijk, om bij het versturen van de gegevens naar de lasermodulator ook de intensiteit van de straal op de juiste momenten te moduleren. De elektronica in de ASIC werk volledig in real time, zodat het afdrukken van een pagina niet vertraagd wordt.

PQET van IBM
Ook IBM heeft in haar Lexmark printers een zélf ontwikkelde PET-technologie toegepast. Nu heet deze "PQET", de afkorting van "Print Quality Enhancement Technology". In feite werkt deze hetzelfde als de RET van Hewlett Packard. Alleen de beeldvlakken die onderzocht worden zijn iets anders. Het imaginaire elektronische beeld dat is opgebouwd in het geheugen van de printer wordt opgedeeld in vlakjes van 9 bij 9 dot's. De samenstelling van deze 81 dot's wordt vergeleken met een groot aantal tabellen, die in het ROM-gedeelte van het geheugen zijn geprogrammeerd. Is de samenstelling van het blok terug te vinden in de tabellen, dan weet de software dat een vaak voorkomende vorm is gedetecteerd en weet het systeem ook hoe de laser gemoduleerd moet worden om deze vorm beter op te vullen.

Opmerking
Alle PET-technieken gaan uit, dat zal na de bespreking van de drie voorbeelden wel duidelijk zijn, van een vorm van patroonherkenning. Als er echter geen patroon te herkennen valt, dan staan deze technieken machteloos. Hetgeen betekent dat PET heel bruikbaar is voor het verbeteren van het resultaat bij het printen van letters en geometrische figuren zoals lijnen en cirkels, maar niets kan doen als er foto's moeten afgedrukt worden. Een foto is immers opgebouwd uit tamelijk willekeurig opeenvolgingen van rastercellen (zie verder) en daar valt weinig patroon in te ontdekken. Een met PET uitgeruste 300 dpi laserprinter zal dus weliswaar een aanzienlijke verbetering te zien geven bij het afdrukken van teksten, maar print foto's nog steeds even onacceptabel af als zijn 300 dpi broertje dat niet "opgewaardeerd" werd!

Het printen van zwart/wit-foto's

Inleiding
Tot voor kort was de standaard resolutie van kantoorprinters 300 dpi. Een dergelijke resolutie leent zich goed voor het afdrukken van teksten en lijntekeningen, maar niet voor het afdrukken van zwart/wit-foto's. Nu echter, wordt 600 en

zelfs 800 dpi de standaard. Door het verhogen van de resolutie kan men meer grijstinten simuleren, waardoor zwart/wit-foto's op een acceptabele manier op het papier verschijnen. In figuur 3/19.11-38 is een dramatisch voorbeeld gegeven van deze stelling. Twee afdrukken van een foto, de linker afgedrukt op een printer met een resolutie van 300 dpi, de rechter op een printer met een resolutie van 1.000 dpi.



Figuur 3/19.11-38: Twee laserafdrukken van een foto, de linker met een resolutie van 300 dpi, de rechter met een resolutie van 1.000 dpi.

Digitale foto's afdrukken

Hoe komt men aan digitale foto's? Dat is op dit moment geen enkel probleem. Er zijn tal van CD-ROM's in de handel, waarop honderden foto's staan in Photo-CD formaat. Met geëigende software kan men deze foto's exporteren naar gestandaardiseerde grafische formaten zoals TIF, BMP en PCX. Daarnaast kan men eigen Photo-CD's laten maken, door fotonegatieven in te leveren bij de fotohandel en deze op CD-ROM te laten zetten. Tot slot kan men natuurlijk foto's scannen, waarbij de illustraties als TIF-file ter beschikking staan of met digitale camera's werken.
Alle moderne tekst-, DTP- en grafische programma's ondersteunen de genoemde formaten, zodat in principe niets in de weg staat om deze foto's in lay-out's op te nemen en af te drukken.
Toch valt er het een en ander te schrijven over het afdrukken van foto's. Het printen van gerasterde afbeeldingen is een eigen specialisme, met eigen wetten en beperkingen.

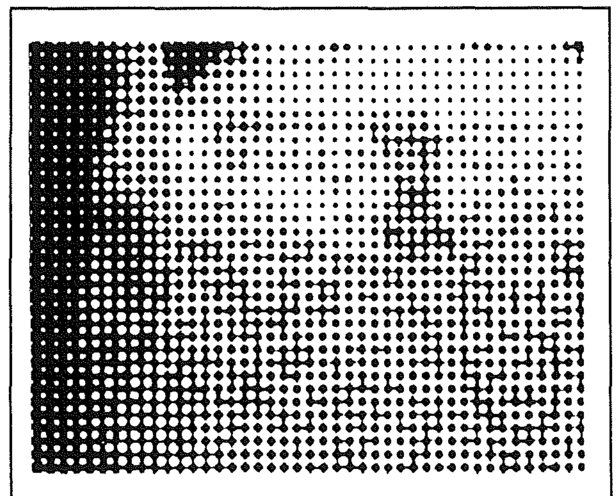
Grijstinten rasteren

Foto's bevatten niet alleen witte en zwarte beelddelen, maar ook grijstinten. Die grijstinten kunnen in principe niet door een printer afgedrukt omdat een printer alleen in staat is dot's zwarte inkt op het papier te zetten. Er moet dus een systeem bedacht worden om de grijstinten van een foto door middel van alleen echt zwarte dot's op het papier te zetten.
De techniek die daarvoor ontwikkeld werd heet "rasteren". Geen systeem dat samenhangt met de moderne digitale beeldverwerking, maar een techniek die reeds meer dan een eeuw wordt toegepast bij kranten- en tijdschriftendruk. Wie een foto in een krant of tijdschrift met een goede loupe bekijkt zal vaststellen dat de op het oog grijze vlakken opgebouwd zijn uit een heleboel kleine, zwarte puntjes. Die puntjes staan zo dicht op elkaar dat het oog alleen maar de gemiddelde zwarting registreert, waardoor in onze hersenen het idee van een grijze tint ontstaat. De mate van grijsheid is niet afhankelijk van het aantal puntjes (dat is voor een bepaalde foto constant over het volledige

oppervlak) maar van de grootte van de puntjes. In figuur 3/19.11-39 is een met de computer gesimuleerde extreme vergroting van een krantenfoto weergegeven. Men merkt duidelijk op dat de puntjes in een horizontaal en vertikaal raster zijn aangebracht en dus op gelijke afstanden van elkaar staan. Alleen de grootte van de puntjes is afhankelijk van de grijstint die weergegeven moet worden. Deze zogenoemde "analoge rasters" worden gemaakt door de originele foto in een reprocamera te belichten op een vel fotogevoelig negatief papier. Op dit vel wordt echter een rasterfolie gelegd. Deze folie is samengesteld uit een grote hoeveelheid minuscule lensjes. Deze lensjes hebben de eigenschap dat zij het invallend licht focusseren op een heel klein puntje van het papier. Als er weinig licht door het lensje valt zal het ontstane puntje klein zijn. Valt er veel licht door het lensje, dat ontstaat een veel groter rasterpuntje. Nadien wordt het negatief papier in een ontwikkelmachine samen met een tweede vel fotogevoelig papier door een chemisch bad gevoerd. De op deze manier geproduceerde gerasterde foto kan dan in een lay-out ingemonteerd worden.
De voornaamste eigenschap van dit procédé is dat de rasterpuntjes een oneindig aantal afmetingen kunnen hebben tussen een minimale waarde en een maximale waarde. Het analoge rasterprocédé is dus in principe in staat een oneindig aantal verschillende grijstinten om te zetten in even zovele rasterdichtheden.

De rasterfrequentie

Het zal duidelijk zijn dat de optisch waargenomen fijnheid van een gerasterde foto afhankelijk is van het aantal rasterpuntjes waaruit het beeld is samengesteld. Hoe meer puntjes, hoe fijner het resultaat. Deze eigenschap wordt de rasterfrequentie genoemd en uitgedrukt door het aantal rasterlijnen per inch. Zoals uit figuur 3/19.11-39 duidelijk blijkt, staan de rasterpuntjes keurig op lijnen. Het aantal lijnen met puntjes dat per inch aanwezig is, wordt gebruikt als definitie voor de fijnheid van het rasterprocédé. De eenheid is lpi, lines per inch, hoewel in Europa soms ook wel de eenheid lpc, lijnen per centimeter, wordt gebruikt. Krantenfoto's worden meestal gerasterd met 85 lpi, foto's in "glossy" magazines gaan tot 120 en zelfs 133 lpi.



Figuur 3/19.11-39: Een (gesimuleerd) vergroot raster van een krantenfoto.

De rasterhoek

Een andere eigenschap van gerasterde afbeeldingen is de zogenoemde "rasterhoek". In figuur 3/19.11-39 lopen de rasterlijnen horizontaal. Dat hoeft niet, vaak is het zelfs beter de rasterlijnen onder een hoek van 45° te laten lopen. Waarom dat beter is heeft iets te maken met de manier waarop het te bedrukken papier door de printer wordt gevoerd. Ontstaan bij dit transport zeer kleine snelheidsafwijkingen, dan worden de in principe ronde

19.11 Principes en werking van printers

rasterpuntjes vertikaal of horizontaal "uitgesmeerd". De puntjes worden enigszins ellipsvormig, met als gevolg een iets donkerder streep door de foto. Rastert men onder een hoek van 45°, dan vallen deze strepen veel minder op.

Dot gain

Een belangrijk verschijnsel waar men bij het verwerken van foto's rekening mee moet houden is de zogenoemde "dot gain" of "punt verdikking". Een gerasterde foto die er op het beeldscherm uitstekend uitziet, kan op de printeruitvoer veel te donker worden. Men zegt dat de afbeelding "dicht loopt". Dit is het gevolg van de manier waarop het papier in de printer wordt behandeld. De rasterpuntjes worden door de laserstraal op de drum belicht, nemen tonerpoeder op en zetten dit elektrostatisch over op het papier. Nadien wordt dit poeder in het papier gebrand. Bij beide stappen worden de puntjes iets uitgesmeerd, zodat gerasterde foto's veel zwarter op het papier zullen verschijnen dan de bedoeling was. Dit verschijnsel noemt men "dot gain", letterlijk vertaald "punt winst" of meer gebruikelijk "punt verdikking". De enige oplossing om dit probleem te verhelpen is de afbeelding voorbereiden in een goed beeldbewerkingsprogramma, door de helderheid en eventueel het contrast van het beeld te manipuleren. Standaard richtlijnen zijn hiervoor niet te geven. Het is een kwestie van rekening houden met het soort papier waarop geprint wordt, het printerprocédé, experimenteren, ervaring opdoen en inschatten hoe een gerasterde foto uiteindelijk op het papier zal verschijnen.

Natuurlijk moet men er op letten dat door het manipuleren van intensiteit en contrast niet het andere uiterste gebeurt, namelijk dat de lichtgrijze puntjes helemaal wit worden. Na het rasteren van de afbeelding door de printer zal er dan helemaal geen contrast in de lichtste beeldpartijen aanwezig zijn. Men noemt dit "het weg lopen" van het beeld. Men moet dus een compromis zoeken tussen enerzijds voorkomen dat de donkere partijen dicht lopen en anderzijds voorkomen dat de lichtste partijen weg lopen.

Beeldscherm versus printer

Bij het verwerken van halftoon afbeeldingen met behulp van computer en printer doet zich een aantal problemen voor. Bij het scannen van een foto wordt het beeld omgezet in een pixelmatrix van bijvoorbeeld 300 bij 300 puntjes per vierkante inch (resolutie 300 dpi). Dit is nog geen raster, maar de best mogelijke representatie van het beeld door middel van puntjes waarvan men de intensiteit op het beeldscherm kan variëren. Als men scant met 8 bit, dan wordt de intensiteit of zwarting van ieder puntje voorgesteld door een getal tussen 000 en 256. Code "000" staat dan bijvoorbeeld voor helemaal wit, code "256" voor helemaal zwart. Daartussen kan de intensiteit van het beeld gedefinieerd worden door 254 grijswaarden. Een "normale" laserprinter met een resolutie van 300 dpi kan echter maar 300 maal 300 identieke en even zwarte puntjes per vierkante inch op het papier zetten. Het zal dus duidelijk zijn dat de 300 x 300 puntjes per vierkante gescande inch, ieder met hun eigen grijswaarde, nooit als dusdanig afgedrukt kunnen worden! De laserprinter kan immers geen grijzen produceren.

Digitaal rasteren

Grijstinten kunnen op een laserprinter alleen gesimuleerd worden door digitaal

19.11 Principes en werking van printers

te rasteren. De laserprinter maakt grijswaarden door in matrixen van bijvoorbeeld 4 bij 4 of 8 bij 8 printposities sommige puntjes wel af te drukken en andere niet. Dit noemt men de "rastercel". Hierdoor ontstaat een patroon, waarbij de gemiddelde grijswaarde afhankelijk is van de hoeveelheid puntjes die worden afgedrukt. Het zal duidelijk zijn dat ook nu ons oog op een schaal van 300 puntjes per inch de afzonderlijke puntjes niet meer waarneemt maar, net zoals bij analoog rasteren, de gemiddelde grijsindruk naar de hersenen stuurt.

Gevolgen voor de afdrukresolutie

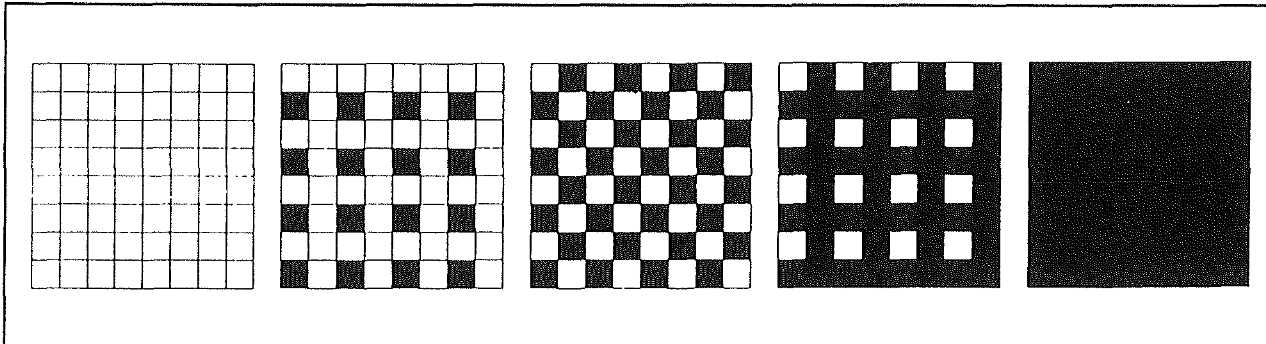
Het principe van digitaal rasteren, grijstinten simuleren door in rastercellen van x bij y puntjes sommige puntjes wel en andere niet af te drukken, heeft dramatische gevolgen! Stel dat een gescande grijstint afbeelding met een resolutie van 300 dpi moet worden afgedrukt op een laserprinter met dezelfde resolutie. De grijstinten in de afbeelding kunnen alleen op de laserprinter gesimuleerd worden door een vermindering van de afdrukresolutie. De software berekent in het eenvoudigste geval de gemiddelde grijswaarde van vier naast elkaar gelegen pixels (gelijk aan een rastercel van 2 bij 2) en stuurt de printer dusdanig aan dat die een van de vijf in figuur 3/19.11-40 getekende patroontjes op het papier print. Tussen "volledig wit" en "volledig zwart" staan dus slechts drie grijswaarden ter beschikking!
De hoge resolutie scan met 256 grijswaarden verschijnt op het papier met een gereduceerde resolutie van slechts 150 dpi en met slechts vijf grijswaarden, inclusief wit en zwart! Van de mooie foto zal dus weinig op het papier overblijven.
Natuurlijk kan men het aantal afdrukbare tinten opvoeren door de afmetingen van de rastercellen te vergroten. In figuur 3/19.11-41 is bijvoorbeeld een rastercel van acht bij acht laserdot's getekend, waarmee men in staat is in totaal 17 grijstinten op het papier te drukken. Maar van de resolutie van 300 bij 300 dpi van de scan blijft op papier dan een resolutie van slechts 300 gedeeld door 8 oftewel 37,5 lpi over, een kwaliteit die niet eens bevredigend is voor goedkoop krantendrukwerk. Natuurlijk is het probleem in de praktijk niet zo dramatisch als bij dit voorbeeld weergegeven.
In de tekening van figuur 3/19.11-41 zijn voor de eenvoud slechts 17 invulpatronen van de rastercel getekend. Met een 6 bij 6 of 8 bij 8 rastercel kan men echter veel meer patronen afdrukken, zodat het aantal afdrukbare grijstinten in de praktijk hoger ligt. Een en ander is natuurlijk volledig afhankelijk van de printersoftware. Algemeen gesteld moet het mogelijk zijn met een 300 dpi laserprinter 33 grijswaarden af te drukken bij een rasterfrequentie van 53 lpi. Nog steeds veel slechter dan de 85 lpi van kranten en met een onbevredigende grijsweergave, maar in ieder geval bruikbaar voor intern gebruik.

Het verband tussen grijswaarden, printerresolutie en rasterfrequentie

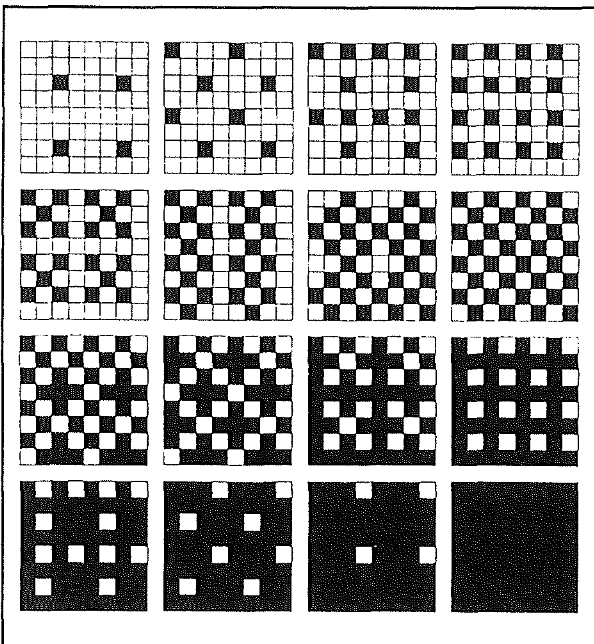
Tussen het aantal af te drukken grijswaarden, de resolutie van de printer en het gewenste aantal lpi van de rasterfrequentie bestaat een eenvoudig wiskundig verband.
Het maximaal aantal grijswaarden is gelijk aan het kwadraat van de verhouding tussen printerresolutie en rasterfrequentie plus één:

$$\text{grijswaarde} = \left\{\frac{\text{printerresolutie}}{\text{rasterfrequentie}}\right\}^2 + 1$$

19.11 Principes en werking van printers



Figuur 3/19.11-40: Het principe van digitaal rasteren met een rastercel van twee bij twee laserpuntjes.



Figuur 3/19.11-41: Als de rastercel uit 8 bij 8 laserpunten bestaat, kan men bijvoorbeeld 17 grijswaarden (inclusief wit en zwart) simuleren.

Met deze formule kan men berekenen dat een 300 dpi printer die een halftoon af moet drukken met 53 lpi maximaal 33 grijstinten kan weergeven. Een 1.000 dpi laserprinter die een foto met 100 lpi bewerkt, kan 101 grijstinten op het papier zetten.

Beperkingen van het oog

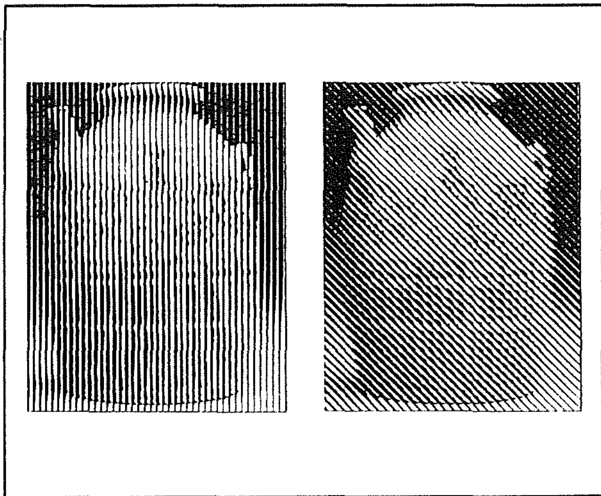
Uit onderzoeken is gebleken dat het menselijke oog niet erg gevoelig is voor kleine verschillen in de grijswaarden van een beeld. Afhankelijk van de eigenschappen van het oog en een eventuele training kan een mens tussen 80 en 120 helderheidsvariaties onderscheiden. Dit maakt het afdrukprobleem natuurlijk minder erg. Het is nergens voor nodig te streven naar een gerasterde afdruk die 256 grijswaarden bevat. Een raster met 100 grijswaarden is goed genoeg. Bij lage resolutie printers kan men dus de gegeven formule omrekenen naar de noodzakelijke rasterfrequentie om een bevredigend aantal grijswaarden (stel 100) op papier te zetten. De minimale rasterfrequentie kan dan (afgerond) berekend worden door de resolutie van de printer te delen door de vierkantswortel van het aantal grijstinten, dus te delen door tien. Een 1.000 dpi laserprinter die 100 grijstinten moet kunnen weergeven kan deze afbeelding met 100 lpi op het scherm zetten.

Printers met "enhancement"-technieken

Bij deze bespreking is tot nu toe uitgegaan van "standaard" laserprinters.
Printers die alleen in staat zijn even grote puntjes toner op het papier te zetten. Sommige fabrikanten hebben echter zeer

19.11 Principes en werking van printers

ingenieuze "print enhancement technieken" ontwikkeld om ook het afdrukken van gerasterde afbeeldingen te verbeteren. Apple gebruikt in zijn LaserWriter IIg een techniek die "Phote Grade" wordt genoemd. Hierdoor is deze 300 dpi printer in staat bij een rasterfrequentie van 106 lpi 67 grijstinten te produceren. Ook het reeds besproken "RET" van Hewlett Packard kan, in de allermodernste versie, halftonen bij een rasterfrequentie van 85 lpi met 101 grijzen afdrukken.



Figuur 3/19.11-42: Door de "screen type" in te stellen op lijnvormig en een zeer lage rasterfrequentie te kiezen (30 lpi) ontstaan speciale effecten.

Gerasterde afbeeldingen printen in de praktijk

Nadat men een af te drukken foto heeft omgezet in bij voorkeur TIF-formaat kan men de foto inlezen in een tekstverwerker met grafische functies of een DTP-pakket. Eerst hier kan men de eigenschappen van het raster definiëren:

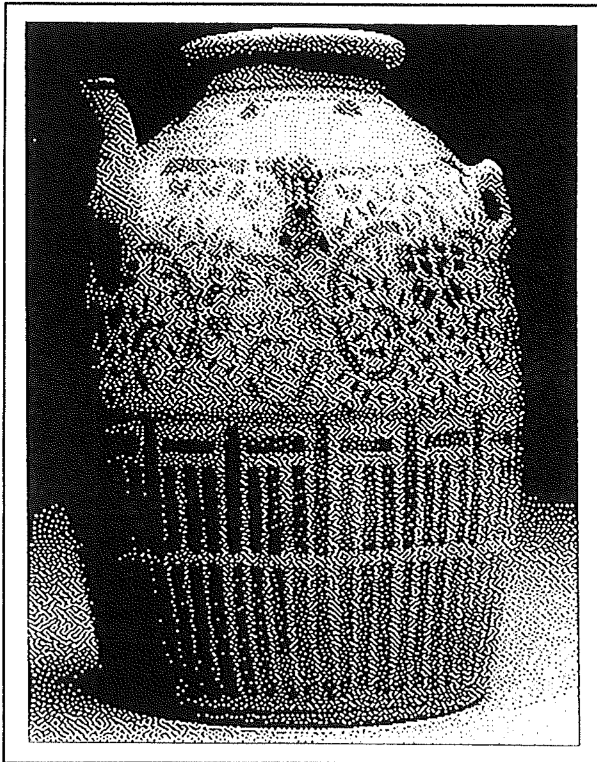
- Screen type
  Met het "screen type" kan men de vorm van de rastercellen definiëren. Deze kunnen rond zijn, ellipsvormig of lijnvormig. Hoewel men meestal voor ronde rastercellen kiest, kan men speciale effecten bereiken door lijnvormige cellen te kiezen en een niet al te hoge rasterfrequentie in te stellen. In de voorbeelden van figuur 3/19.11-42 worden bijvoorbeeld twee lijnvormige rasters voorgesteld, de ene onder een hoek van 90° en de tweede onder een hoek van 45°. Door de rasterfrequentie op een lage waarde in te stellen, 30 lpi in dit geval, ontstaat een speciaal effect dat ook wel eens in kranten wordt toegepast om een bepaalde foto te onderscheiden van de overige.
- Screen Angle
  Hiermee kan men uiteraard de rasterhoek instellen. Standaard waarde is 45°, maar soms ontstaan interessante effecten als men andere hoeken kiest.
- Line Screen
  Hiermee kan men de rasterfrequentie instellen. Het heeft geen enkele zin een hogere rasterfrequentie in te stellen dan de resolutie van de printer toestaat. Het enige gevolg is dat het raster volledig dicht loopt!

Dithering

Wie niet over een hoge resolutie printer beschikt en toch foto's wil afdrukken, zal beroep moeten doen op een techniek die "dithering" wordt genoemd. Dithering is een techniek waarbij grijswaarde illustraties op een speciale manier worden omgezet in vrij groffe zwart/wit-afbeeldingen. Het lijkt op rasteren, maar het heeft er helemaal niets mee te maken.

Bij dithering wordt de grijsinformatie van een illustratie omgezet in bepaalde quasitoevallige combinaties van witte en zwarte dot's. Dithering werkt niet dot per dot, maar berekent de gemiddelde grijswaar-

de van een aantal pixels en vervangt deze grijswaarde door een bepaald zwart/wit-patroon. Bij 300 dpi printers werkt dithering vaak beter dan rastering. Omdat geditherde afbeeldingen alleen zwart/wit-informatie bevatten zijn de files natuurlijk veel kleiner dan deze van gerasterde plaatjes. In figuur 3/19.11-43 wordt als voorbeeld de pot van figuur 3/19.11-42 geditherd weergegeven. Deze TIF-file is slechts 17,8 kB groot, terwijl de 265 grijstinten file 138 kB groot is.



Figuur 3/19.11-43: Een voorbeeld van een geditherde foto, afgedrukt op een laserprinter met een resolutie van 300 dpi.

Soorten dithering

De grote vraag die beantwoord moet worden is hoe het mogelijk is dat met een dergelijke zeer groffe zwart/wit-benadering de indruk van een grijsbeeld zo goed benaderd kan worden. Dat is een kwestie van wiskunde en dus van bepaalde algoritmen. Dithering kan uitgevoerd worden met verschillende algoritmen, die de naam hebben gekregen van hun ontdekker:

- Bayer
  Dit is het eenvoudigst en snelst werkend algoritme, dat uitstekend bruikbaar is om snel een indruk te krijgen van de inhoud van halftoon illustraties. De meeste grafische programma's die zogenoemde "thumbnails" (kleine afbeeldingen die een overzicht op het scherm geven van alle grafische files op een harde schijf) aanmaken werken met dit algoritme.
- Floyd-Steinberg
  Dit algoritme werkt volgens het wiskundige principe van "fout-diffusie". Het levert heel bruikbare resultaten op, maar werkt vrij traag. Dat wil zeggen dat het ditheren van een halftoon scan vrij veel tijd kost, zelfs op snelle Pentium-computers. Het voorbeeld van figuur 3/19.11-43 is volgens dit algoritme geditherd.
- Stucki
  Dit is het allerbeste algoritme en het werkt ook volgens het "fout-diffusie"-procédé. Bepaalde vereenvoudigingen die "Floyd-Steinberg" vanwege de snelheid invoert worden echter niet doorgevoerd. Ditheren volgens "Stucki" gaat dus het traagst, maar zou de beste resultaten moeten opleveren. Maar de waarheid gebiedt de stellen dat men nauwelijks verschillen opmerkt tussen een met "Stucki" of een met "Floyd-Steinberg" geditherde file.
- Burkes
  Het "Burkes"-algoritme ligt ergens tussen de twee eerder beschreven "fout-

diffusie"-systemen in en zou snelheid met kwaliteit moeten combineren. Maar alweer, zou men de pot ditheren volgens Burkes, dan zou men nauwelijks verschil zien met figuur 3/19.11-43.

Besluit
Samengevat kan men stellen dat dithering een interessante optie is als men foto's wil afdrukken op een printer die een lage resolutie heeft. Alle inkjet-printers werken op dit moment bijvoorbeeld volgens dit procédé. Door het "spatten" van de inkt zijn de hedendaagse inkjet-printers nog niet in staat echte rastercellen op papier af te drukken. Wie bovendien niet meer heeft dan een oude matrixprinter kan met dithering redelijke afdrukken krijgen van grijswaarden plaatjes. Voor dit soort printers is dithering de enige manier om halftonen af te drukken!

Het printen van kleuren-foto's

Problemen bij kleurgebruik
Wie voor het eerst ontwerpen in kleur afdrukt op een kleurenprinter, zal vaak constateren dat de resultaten op papier tamelijk teleurstellend zijn. Een ontwerp, dat er op het beeldscherm schitterend uitziet, komt vaal en bleek uit de printer rollen.
Een bepaalde kleur, zélf gedefinieerd op het scherm, verschijnt als een geheel andere tint op het papier. Dergelijke problemen zijn niet zo verbazingwekkend, want bij het afdrukken in kleur komt heel wat kijken, veel meer dan de leek denkt! Het grootste probleem is dat de kleuropbouw van een monitor heel anders werkt dan de kleuropbouw van een kleurenprinter. Wie dus professioneel met kleur aan de slag wil, zal eerst wat beginselen van de kleurentheorie moeten kennen.

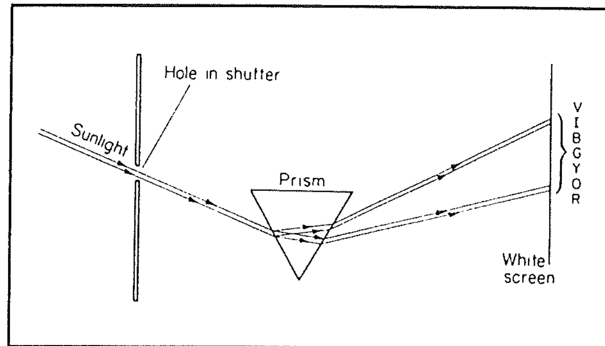
Het licht, de basis van kleur
Kleuren zijn alleen waarneembaar als er licht is. Licht is een vorm van energie, die zich uit onder de vorm van elektromagnetische straling.
Volgens de klassieke mechanica plant die straling zich voort door de ruimte als een elektromagnetisch golfverschijnsel. Golven zijn een algemeen bekend begrip, denk maar aan de golven die door de zee op het strand worden geworpen. Het fundamentele verschil tussen deze watergolven en de elektromagnetische golven van het licht is dat de eerste soort "materieel" aanwezig is, dus voelbaar, terwijl de tweede soort niet tastbaar is, maar door levende wezens alleen met de ogen waar te nemen is als licht.
Golven hebben bepaalde eigenschappen, waarvan de voornaamste de golflengte is. Iedereen kent dit verschijnsel, immers watergolven hebben toppen en dalen. Het ligt voor de hand dat de golflengte van een golf de afstand is tussen twee opeenvolgende toppen of twee opeenvolgende dalen.
De eenheid van golflengte is dan ook de meter. Het probleem bij elektromagnetische golven is dat deze niet als dusdanig zichtbaar zijn en men de golflengte dus niet kan meten met een meetlat. Toch heeft men apparatuur ontwikkeld, waarmee de golflengte van die onzichtbare elektromagnetische golven heel nauwkeurig te meten is. En wat blijkt dan? Dat de grootte van de golflengte rechtstreeks verband houdt met de kleur die onze ogen waarnemen!

Kleur en golflengte

Elektromagnetische golven met een golflengte van 700 nm (afkorting van 700 nanometer, is gelijk aan 0,000.000.7 meter!) die in onze ogen binnen vallen, worden ervaren als rood gekleurd licht. Elektromagnetische golven met een golflengte van 360 nm worden door onze ogen ervaren als violet licht. Tussen die twee grenzen ligt het gehele spectrum van waarneembare kleuren, een spectrum dat maar al te goed bekend is van de regenboog.
Hiermee is dus de vraag beantwoord wat kleur is.
Kleur is niets meer of minder dan een menselijke interpretatie van de golflengte van elektromagnetische straling.

Wit licht

Naast alle mogelijke kleuren neemt de mens ook wit licht waar. Wat is wit licht? Wit licht heeft geen specifieke golflengte, maar bestaat uit een menging van elektromagnetische golven met ieder hun eigen specifieke golflengte. Wit is dus een mengkleur, een eigenschap die eeuwen geleden door de Engelse natuurkundige Newton werd ontdekt. Newton stuurde het wit licht van de zon door een prisma, een driehoekvormig stuk glas. Glas heeft de eigenschap dat het elektromagnetische golven kan afbuigen, dus van richting kan laten veranderen. De mate van afbuiging is echter recht evenredig met de golflengte van de straling.
Het gevolg van een en ander is dat alle golflengten, die in wit licht aanwezig zijn, onder een andere hoek worden afgebogen en dus aan de andere kant van het prisma gescheiden uittreden. Stuurt men wit licht door een prisma, dan ontstaat aan de andere kant van het prisma een soort regenboog, die alle waarneembare kleuren bevat van donker rood tot donker paars, zie figuur 3/19.11-44.



Figuur 3/19.11-44: Het experiment van Newton, waarbij het wit licht van de zon werd ontleed in de samenstellende kleuren.

Mengkleuren

Uit het experiment blijkt een belangrijke eigenschap van het menselijk oog. Valt een elektromagnetische golf met een typische golflengte binnen in ons oog, dan interpreteren wij dat als een bepaalde kleur. Dezelfde kleurinterpretatie kan echter ook ontstaan als twee, drie, of meer golven binnen vallen met ieder een specifieke golflengte. In dit geval is er sprake van menging. Stel dat ons oog wordt getroffen door twee elektromagnetische golven, een “groene” en een “rode”. De termen “groen” en “rood” hebben uiteraard betrekking op de golflengten van de straling. Gehoor gevend aan de universele wetten van de natuurkunde, ontstaat er een menggolf met specifieke eigenschappen, die door ons oog geïnterpreteerd worden als zijnde een golf met een “gele” golflengte.

Primaire kleuren

Het feit dat het menselijk oog ook mengingen van verschillende golflengten waarneemt als kleur is een zeer belangrijk

19.11 Principes en werking van printers

gegeven. Zowel bij het produceren van kleuren op een beeldscherm als op een kleurenprinter wordt namelijk dankbaar gebruik gemaakt van dit verschijnsel. Men heeft ontdekt dat er drie primaire kleuren bestaan, waaruit door menging alle waarneembare kleuren kunnen worden samengesteld. Die primaire kleuren zijn rood, groen en blauw. Uit elektromagnetische stralingen met deze drie basis golflengten kunnen miljoenen kleurschakeringen worden samengesteld. Het volstaat de sterkte of intensiteit van de drie primaire kleuren te variëren. Even sterke intensiteiten rood en blauw licht geven een mengkleur, die door ons ervaren wordt als magenta. Voert men de intensiteit van de blauwe straling iets op, dan zal het magenta een blauwachtige verkleuring ondergaan.
Op dezelfde manier kan men wit licht maken, door een welbepaalde combinatie van rode, groene en blauwe straling te genereren, ieder met een specifieke intensiteit.

Mengsystemen

Bij het mengen van kleuren moet echter onderscheid worden gemaakt tussen "additieve" kleurwaarneming en "substractieve" kleurwaarneming. Een zeer belangrijk onderscheid, want een monitor werkt bijvoorbeeld additief, maar een kleurenprinter substractief. Als in de praktijk blijkt dat een geprint plaatje wat kleursamenstelling betreft absoluut niet lijkt op het mooi gekleurd plaatje van het beeldscherm, dan ontstaat dit probleem door het fundamentele verschil tussen additieve en substractieve kleurmenging!
Kleuren, die worden uitgestraald door een lichtbron, zijn ***additieve*** kleuren. Die additieve kleuren kunnen rechtstreeks in ons oog binnenvallen en geven dan een bepaalde kleurindruk. Stel dat men op het podium van een theater staat en dat er drie spot's op het podium gericht worden, die ieder een van de primaire kleuren uitstralen. Hoe meer licht er door die spot's wordt uitgestraald, hoe lichter de totaalindruk zal worden. De lichttechnicus kan nu de intensiteit van de drie spot's zelfs zo instellen, dat de indruk van wit licht ontstaat.
Additieve kleuren ontstaan dus door *lichtenergie op te wekken.* Beeldschermen werken additief, omdat zij licht uitzenden.
Substractieve kleuren ontstaan door *lichtenergie te onttrekken.* Dit is een natuurkundig verschijnsel dat enige toelichting behoeft. Een citroen zal altijd als geel worden ervaren, ondanks het feit dat deze vrucht absoluut geen lichtbron is. Hoe ontstaat dan deze gele kleur? Simpelweg, door reflectie. De citroen wordt door een lichtbron, bijvoorbeeld de zon, beschenen. De moleculen waaruit de schil van de citroen bestaat, hebben als eigenschap dat zij bepaalde golflengten kunnen absorberen en andere reflecteren. Bij een citroen is het nu toevallig zo geregeld dat alleen licht dat een "gele" golflengte heeft wordt gereflecteerd en alle andere golflengten worden geabsorbeerd. Alleen het gele licht bereikt dus via reflectie ons oog, met als logisch gevolg dat het lijkt alsof deze citroen geel is.
Gekleurde vellen papier, zoals zij door een kleurenprinter worden geproduceerd, werken dus substractief. Zie zenden immers geen licht uit, maar reflecteren alleen bepaalde golflengten van het licht dat er op invalt. Hoe meer gekleurde inkt er op het papier aanwezig is, hoe meer invallend licht er wordt geabsorbeerd en hoe minder er wordt gereflecteerd. Hoe meer inkt, hoe donkerder het beeld wordt!

Secundaire kleuren

Als gevolg van het fundamentele verschil tussen de productie van kleuren bij een additief en bij een substractief systeem, werken substractieve systemen nooit ofte nimmer met de drie basiskleuren rood, groen en blauw. Het grote verschil is immers dat een monitorscherm zwart is en een vel papier wit! Zou men op een wit vel papier even veel van de drie primaire kleuren drukken, dan zou het resultaat nog steeds wit zijn! Vandaar dat kleurenprinters, of die nu met drukinkten, toners, pigmenten of verven werken, allemaal zonder uitzondering gebruik maken van de zogenoemde "secundaire" kleuren. Dat zijn er ook drie, te weten:

- Cyaan
  Cyaan is licht blauw en wordt samengesteld uit de primaire kleuren blauw en groen. Cyaan absorbeert rood licht.
- Magenta
  Magenta is vaal rood, wordt samengesteld uit de primaire kleuren rood en blauw en absorbeert alles dat groen is.
- Geel
  Geel is gewoon geel, wordt samengesteld uit de primaire kleuren rood en groen en absorbeert blauw.

Eigenschappen van kleuren

Kleuren worden gedefinieerd door drie belangrijke eigenschappen:

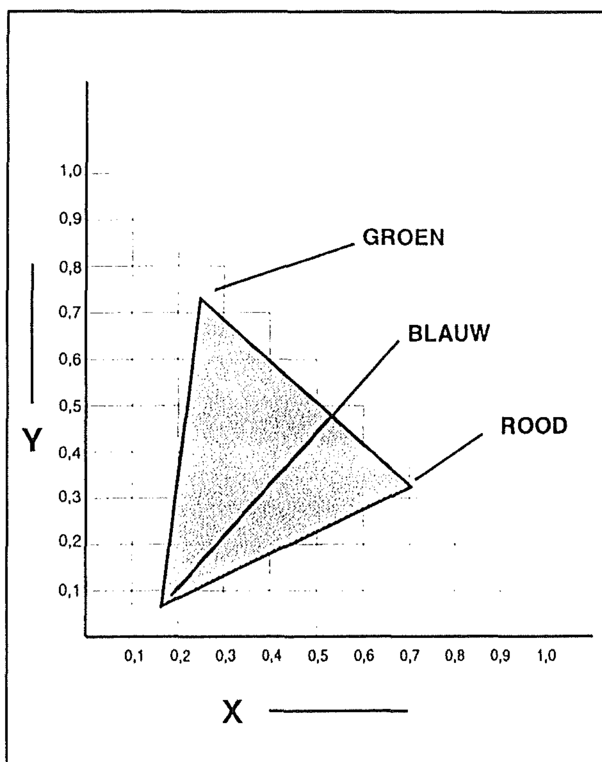
- De tint
  De tint kan het best beschreven worden als de voornaamste primaire kleur die in een mengkleur aanwezig is. De tint van vleeskleur is bijvoorbeeld rood, de tint van de lucht is blauw. Maar de tint van roze, paars of bruin is ook rood. Natuurkundig uitgedrukt is de tint de overheersende golflengte van een van de drie primaire kleuren die in de mengkleur aanwezig is. In het Engels wordt de tint aangegeven door het begrip "Hue".
- De helderheid
  De helderheid kan het best omschreven worden als de donkerte van een bepaalde kleur. Licht groen heeft een grotere helderheid dan donker groen, hoewel in beide gevallen de primaire kleur groen overheersend is. Natuurkundig bekeken kan helderheid worden uitgedrukt als de intensiteit van de overheersende golflengte in de mengkleur. In het Engels wordt de helderheid van een kleur "Brightness" genoemd.
- De verzadiging
  De kleurverzadiging is een begrip dat heel goed is waar te nemen, maar heel moeilijk te omschrijven is. Een bepaalde kleurtint kan sprankelend en levendig zijn, maar ook onopvallend grijzig. Het zal duidelijk zijn dat de verzadiging heel veel met de helderheid te maken heeft. Natuurkundig gesproken bevat een verzadigde kleur heel veel straling van een van de drie primaire kleuren en heel weinig straling van de twee overige primaire kleuren. In het Engels wordt deze eigenschap van kleur "Saturation" genoemd.

Het definiëren van kleuren

Moderne monitoren en kleurenprinters kunnen niet minder dan 16.777.216 verschillende kleuren genereren. Hoe nu al deze kleuren van elkaar te onderscheiden? Het zal duidelijk zijn dat het weinig zin heeft namen toe te kennen. "Citroengeel" of "Komkommergroen" zijn te wazige begrippen om er mee te kunnen werken. Vandaar dat er internationaal een aantal kleurensystemen is ontwikkeld, waarvan:

- de kleurendriehoek volgens DIN 5033;

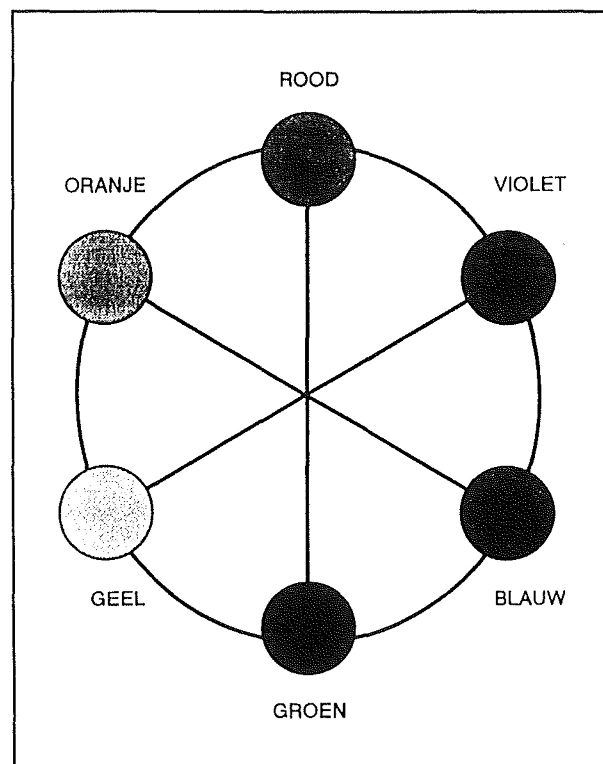
– de kleurencirkel;
– het "PMS"-systeem;
wel de meest bekende zijn.



Figuur 3/19.11-45: De volgens DIN 5033 genormeerde internationale kleurendriehoek.

De kleurendriehoek

Uit de drie primaire kleuren rood, groen en blauw kunnen, zoals reeds beschreven, door additieve menging miljoenen kleursoorten samengesteld worden. Al die kleuren kunnen overzichtelijk worden samengevat in "de kleurendriehoek". Zoals getekend in figuur 3/19.11-45 bestaat die driehoek uit een assenstelsel met twee coördinaten X en Y. Dit kleurcoördinatensysteem is genormd volgens DIN 5033 en wordt internationaal toegepast. De drie primaire kleuren R, G en B liggen op de hoekpunten van de driehoek. Langs de zijden van de driehoek gaat de kleur geleidelijk over van de ene kleur in de andere. Tussen de rode en de groene hoek wisselen de kleuren, net zoals bij een regenboog, van rood via geel naar groen. Iedere kleur kan nu gemakkelijk gedefinieerd worden door er een X- en een Y-waarde aan toe te kennen. De primaire kleur groen heeft bijvoorbeeld een X-waarde van ongeveer 0,2 en een Y-waarde van ongeveer 0,7. Het zogenoemde "witpunt", het punt waar uit een bepaalde menging van de drie primaire kleuren wit ontstaat, ligt bij X = 0,3 en Y = 0,3.



Figuur 3/19.11-46: De kleurencirkel, die gemakkelijk kan afgeleid worden uit de kleurendriehoek.

De kleurencirkel

Uit de kleurendriehoek van figuur 3/19.11-46 kan gemakkelijk de kleurencirkel van figuur 3/19.11-45 afgeleid worden. Men kan zich voorstellen dat de kleurendriehoek R-G-B uit een gekleurd snoer bestaat dat rond drie spijkers is gespan-

nen. Verwijdert men de spijkers en legt men het snoer in een cirkel, dan ontstaat de kleurencirkel.

Het PMS-systeem

PMS is de afkorting van "Pantone Matching System". Het PMS-systeem bestaat uit een kleurenwaaier die duizenden kleuren bevat. Iedere kleur heeft een bepaald nummer. Van iedere kleur is bovendien tot op de procent nauwkeurig gedefinieerd hoe die kan worden samengesteld uit de primaire kleuren rood, groen en blauw of uit de secundaire kleuren cyaan, magenta en geel. Het PMS-systeem definieert dus zowel het additieve als het substractieve mengproces. Betere grafische programma's, zoals Corel Draw, hebben de mogelijkheid om kleuren samen te stellen volgens het PMS-systeem. Wil men een bepaald vlak op het scherm vullen met een PMS-kleur, dan volstaat het dit element van het ontwerp aan te klikken, een venster te openen en de PMS kleurencode in te vullen. Het programma stuurt dan zoveel rood, groen en blauw naar het scherm, dat de gewenste tint ontstaat. Hetzelfde geldt nadien bij het printen van deze kleur. Goede printerdrivers van kleurenprinters zijn in staat de PMS-code van ieder kleur te achterhalen en de noodzakelijke combinatie van cyaan, magenta en geel op het papier te spuiten.

Scherm/papier compatibiliteit

Die PMS-compatibiliteit vormt het grootste probleem bij het ontwerpen van documenten die kleuren bevatten. Als alle gebruikte software PMS-compatibel is, zijn er minimale problemen te verwachten. Hoewel, de ijking van het beeldscherm en de eigenschappen van het papier spelen zelfs dan een belangrijke rol (zie later)! Als echter het tekenprogramma kleuren op een eigen manier samenstelt is het maar helemaal de vraag of de driver van de kleurenprinter die eigen codering kan ontcijferen. Het gevolg kan dan zijn dat iets dat licht groen op het beeldscherm verschijnt, licht paars wordt afgedrukt.

Complementaire kleuren

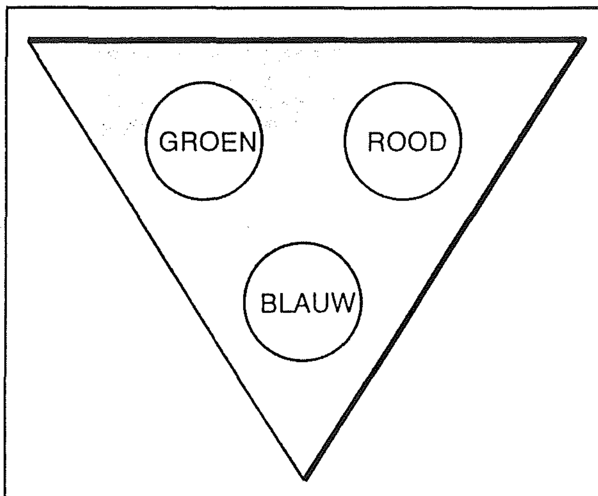
Tot slot van deze korte introductie in de kleurenleer moet nog even aandacht worden besteed aan een belangrijk begrip, namelijk complementaire kleuren. Complementaire kleuren zijn kleuren die zich op de kleurencirkel tegenover elkaar bevinden. Complementaire kleuren zijn dus rood en groen of oranje en blauw. Het menselijk oog heeft iets met complementaire kleuren. Uit ervaring blijkt bijvoorbeeld dat het heel onverstandig is complementaire kleuren in een en dezelfde lay-out te verwerken!

Kleuren printen in de praktijk

Het RGB-systeem van de monitor

Zoals reeds geschreven, werkt een monitor additief. Er moeten dus lichtbronnen aanwezig zijn, die licht uitstralen. Die lichtbronnen zijn aanwezig onder de vorm van zogenoemde pixeldriehoeken. Dat zijn uiterst kleine driehoekjes, die ieder weer zijn opgebouwd uit drie fosforpuntjes. Deze puntjes zijn een fractie van een millimeter groot! De elektronica in de beeldbuis straalt elektronen naar die fosforpuntjes. Als een fosforpuntje door elektronen wordt getroffen, zal dit puntje licht uitstralen. Bij een kleurenbeeldbuis worden de fosformaterialen zo gekozen, dat één puntje rood, één puntje blauw en

één puntje groen licht uitstraalt. Een en ander is voorgesteld in figuur 3/19.11-47. Met zegt, vanwege deze kleuropbouw, dat een monitor volgens het RGB-systeem werkt, de afkorting van rood, groen en blauw.



Figuur 3/19.11-47: De fundamentele kleurencel van een monitor bestaat uit drie uiterst kleine fosforpuntjes.

Kleuren mengen op het scherm

Het volledig scherm van de monitor bevat tienduizenden van deze pixeldriehoeken. Als de elektronica alleen de groene fosforpuntjes aanstuurt, zal het gehele scherm groen oplichten. Worden de drie kleurkanalen digitaal aangestuurd, dus "uit" of "aan", dan kan men met dit RGB-systeem acht kleuren produceren:

R = uit	G = uit	B = uit	zwart
R = aan	G = uit	B = uit	rood
R = uit	G = aan	B = uit	groen
R = uit	G = uit	B = aan	blauw
R = aan	G = aan	B = uit	geel
R = aan	G = uit	B = aan	magenta
R = uit	G = aan	B = aan	cyaan
R = aan	G = aan	B = aan	wit

Moderne grafische kaarten kunnen echter de drie kanalen in intensiteit sturen en wel in 256 verschillende stappen. Dat betekent dat het systeem in staat is in totaal 256 x 256 x 256 = 16.777.216 verschillende kleurentinten op het scherm te zetten.

Het CYMK-systeem van de printer

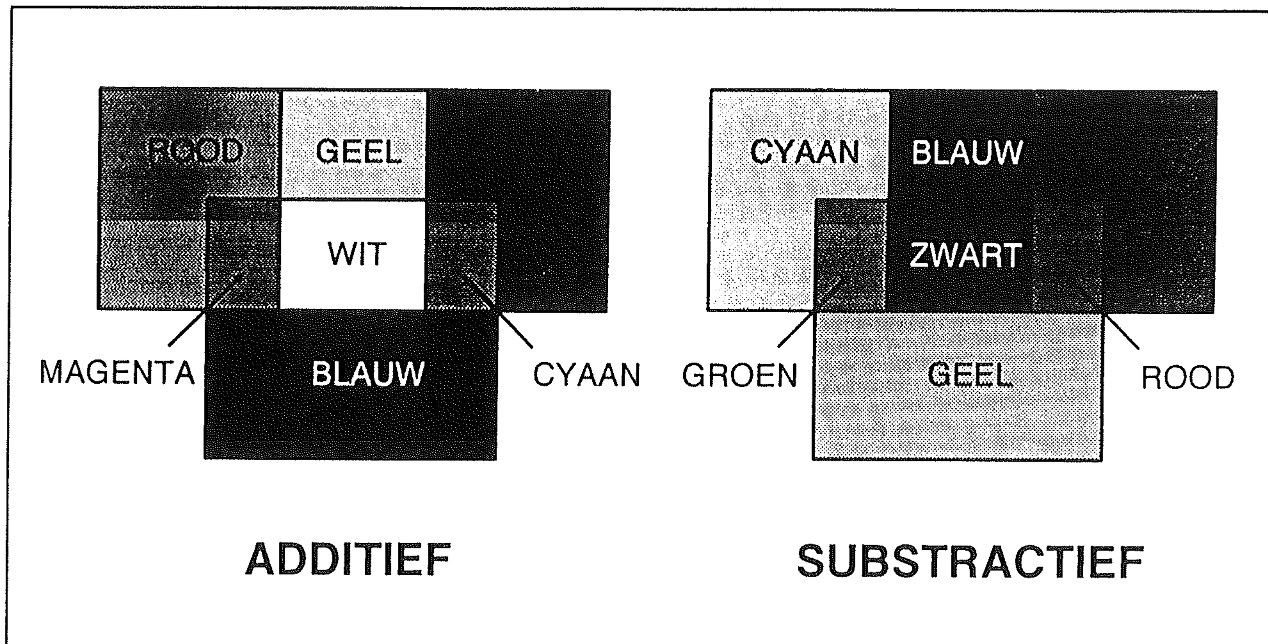
Het kleurprocédé van een printer werkt geheel anders, namelijk volgens het substractieve principe. Papier straalt natuurlijk geen licht uit. Het witte omgevingslicht valt op het papier, dringt door de transparante inkten heen en wordt nadien teruggekaatst door het papier. Bepaalde golflengten worden hierbij geabsorbeerd, andere doorgelaten. Moderne kleurenprinters beschikken over vier inktpatronen, waarvan er drie meestal in één inktcartridge verwerkt zitten:

- cyaan (C);
- geel (Y van Yellow);
- magenta (M);
- zwart (K van blacK).

Het mechanisme van de printer stuurt de constructie, waarin de inktpatronen gemonteerd zijn, heen en weer over het papier. Via de nozzles kunnen uiterst kleine druppeltjes inkt op het papier gespoten worden. Op deze manier kunnen weer acht verschillende kleuren geproduceerd worden. Spuit men namelijk op een en dezelfde plek twee kleuren, dan zullen deze kleuren zich mengen tot een mengkleur. Dus:

C = uit	Y = uit	M = uit	K = uit	wit
C = aan	Y = uit	M = uit	K = uit	cyaan
C = uit	Y = aan	M = uit	K = uit	geel
C = uit	Y = uit	M = aan	K = uit	magenta
C = uit	Y = uit	M = uit	K = aan	zwart
C = aan	Y = aan	M = uit	K = uit	groen
C = aan	Y = uit	M = aan	K = uit	blauw
C = uit	Y = aan	M = aan	K = uit	rood

19.11 Principes en werking van printers



Figuur 3/19.11-48: De mengschema's van beeldscherm (links) en kleurenprinter (rechts) met elkaar vergeleken.

Merk op dat de zwarte inkt alleen wordt gebruikt voor het produceren van zwart. Als men cyaan, magenta en yellow over elkaar heen spuit, ontstaat namelijk geen zwart, maar een zeer donker bruine tint, die een erg smerig zwart oplevert. Vandaar dat men, noodgedwongen, een vierde inktpatroon moet gebruiken.

Vergelijking

In figuur 3/19.11-48 worden de mengschema's van een monitor en een kleurenprinter met elkaar vergeleken.

Rasteren van kleuren

Het zal duidelijk zijn dat gerasterd moet worden om vanuit de vier aanwezige inktpatronen de duizenden mengkleuren te printen die een moderne grafische kaart op het scherm kan zetten. Het toegepaste rasterprocédé is echter afhankelijk van de mogelijkheden van de printer. Bij het rasteren van kleuren moet men onderscheid maken tussen printers die in staat zijn de grootte van de inktdot te moduleren en printers die dat (nog) niet kunnen. Het eerste procédé is op dit moment alleen beschikbaar bij de dure dey sublimation printers. De populaire inktspuiters en laserprinters werken met dotjes van constante grootte.

Omdat de meeste PC-gebruikers op dit moment inkjet- of laserprinters gebruiken, die met vaste inktdotjes werken, worden de rastertechnieken van deze printers besproken.

Men heeft vijf rastersystemen ontwikkeld:

- substractieve dithercellen;
- amplitude gemoduleerd rasteren (AM);
- adaptieve halftoning;
- frequentie gemoduleerd of stochastisch rasteren (FM);
- hybride rasteren.

Er bestaat een fundamenteel verschil tussen het eerste procédé en de overige. Bij toepassing van substractieve rastering wordt gebruik gemaakt van één raster, dat

in één keer op het papier wordt geprint. Bij de vier overige systemen wordt gebruik gemaakt van vier deelrasters, één raster voor iedere inktkleur.

Substractieve dithercellen
Substractief ditheren is een procédé, waarbij kleurentinten worden geproduceerd door diverse basiskleuren aan te brengen in dithercellen. Een dithercel bestaat bijvoorbeeld uit acht bij acht inktpuntjes. Ieder van die 64 inktpuntjes kan nu gevuld worden met een van de acht mogelijke basiskleuren. Bekijkt men een geditherde afdruk met een vergrootglas, dan zijn die afzonderlijke inktpuntjes goed waarneembaar. Op enige afstand echter, vloeien deze kleine inktpuntjes samen en ontstaat de indruk van een mengkleur. In figuur 3/19.11-49 wordt het substractieve dithering-procédé voorgesteld. De dithercel van 8 bij 8 inktpunten wordt gevuld met:
- 2 x zwarte inkt;
- 3 x magenta inkt;
- 3 x cyaan inkt;
- 15 x rode inkt (geel + magenta);
- 41 x gele inkt.

Het resultaat is een vleesachtige kleur.

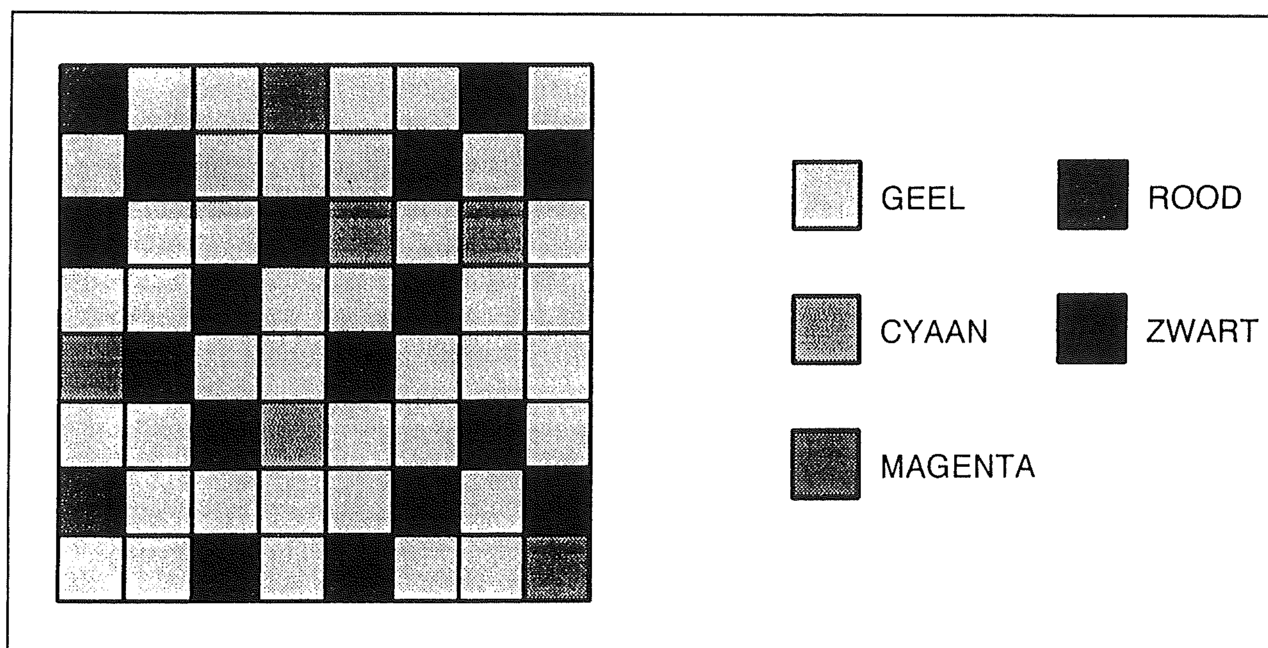
Soorten dithering
Een substractieve dithercel kan op verschillende manieren worden samengesteld. De twee meest toegepaste systemen zijn:
- patroon;
- verstrooid.

Bij "patroon" wordt iedere dithercel die een bepaalde tint moet voorstellen op dezelfde manier met inktpuntjes gevuld. Het nadeel van deze methode is dat vanaf een afstand bekeken regelmatige geometrische patronen kunnen ontstaan, die heel erg storend werken. Bij "verstrooid" worden dithercellen die dezelfde tint moeten voorstellen op een willekeurige manier gevuld met de noodzakelijke kleurpunten. De verdeling naar aantal blijft uiteraard identiek per cel, alleen de plaats waar de puntjes worden geprint verschilt op een willekeurige manier van cel tot cel. De kans op geometrische patronen wordt hierdoor verkleind. Wél ontstaat nu de kans dat in vier naast elkaar gelegen hoeken van vier cellen een opeenhoping van een bepaalde inktkleur ontstaat, waardoor in een groot vlak dat egaal met een bepaalde tint moet worden gevuld plaatselijk kleine kleurafwijkingen kunnen ontstaan. De meeste drivers die bij kleurenprinters worden geleverd, bieden de mogelijkheid de dithering-methode in te stellen op "patroon" of "verstrooid". Welk van beide methoden de beste resultaten geeft moet vaak experimenteel bepaald worden.

Dithering en resolutie
De meeste moderne kleurenprinters hebben een basisresolutie van 360 dpi. Door echter te ditheren gaat de resolutie dramatisch dalen! Iedere dithercel kan immers maar één welbepaalde kleurtint representeren. Als met dus dithercellen van 8 bij 8 inktpuntjes toepast, daalt de resolutie tot 360 gedeeld door 8 is gelijk aan 45 dpi! Vandaar dat men in de meeste gevallen met kleinere dithercellen werkt en met minder reproduceerbare kleuren genoegen neemt.

Amplitude gemoduleerd rasteren
Bij het printen van afbeeldingen in kleur door middel van AM-rastertechnieken doet zich een aantal extra problemen voor. Ook bij AM-rasteren maakt men gebruik van cellen, die uit verschillende inktdotjes bestaan.

19.11 Principes en werking van printers



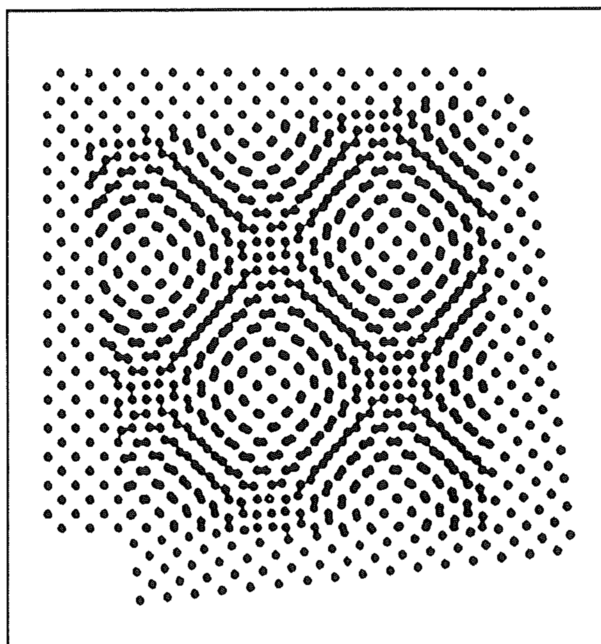
Figuur 3/19.11-49: Door een 8 bij 8 dithercel op de getekende manier te vullen met zwart, geel, cyaan, magenta en rood, ontstaat een huidkleur.

Omdat er nu met vier kleuren wordt gewerkt (cyaan, magenta, geel en zwart) moet men vier keer rasteren. En dat veroorzaakt extra problemen! Het zal duidelijk zijn dat die rasters niet zo maar willekeurig over elkaar heen gedrukt kunnen worden. Het is zelfs absoluut noodzakelijk dat de rasterpuntjes van de vier drukkleuren grotendeels naast elkaar in plaats van over elkaar worden gedrukt.
Een van de grootste uitdagingen bij het AM-rasteren van kleuren in het vermijden van het zogenoemde Moiré-effect. Dit effect wordt toegelicht aan de hand van figuur 3/19.11-50. In deze tekening worden twee rasters onder een willekeurige hoek over elkaar heen gezet.
Zoals duidelijk uit de figuur blijkt, ontstaat er een merkwaardig en zeer regelmatig verlopend geometrisch motief. Dit verschijnsel wordt "het Moiré-effect" genoemd. Het variëren van de hoek tussen de twee rasters met een paar graden kan een heel ander geometrisch patroon tot gevolg hebben! Het zal duidelijk zijn dat men het Moiré-effect in de praktijk kan missen als kiespijn. Vandaar dat men een optimale hoek tussen de vier rasters moet zoeken, waarbij dit effect minimaal optreedt.

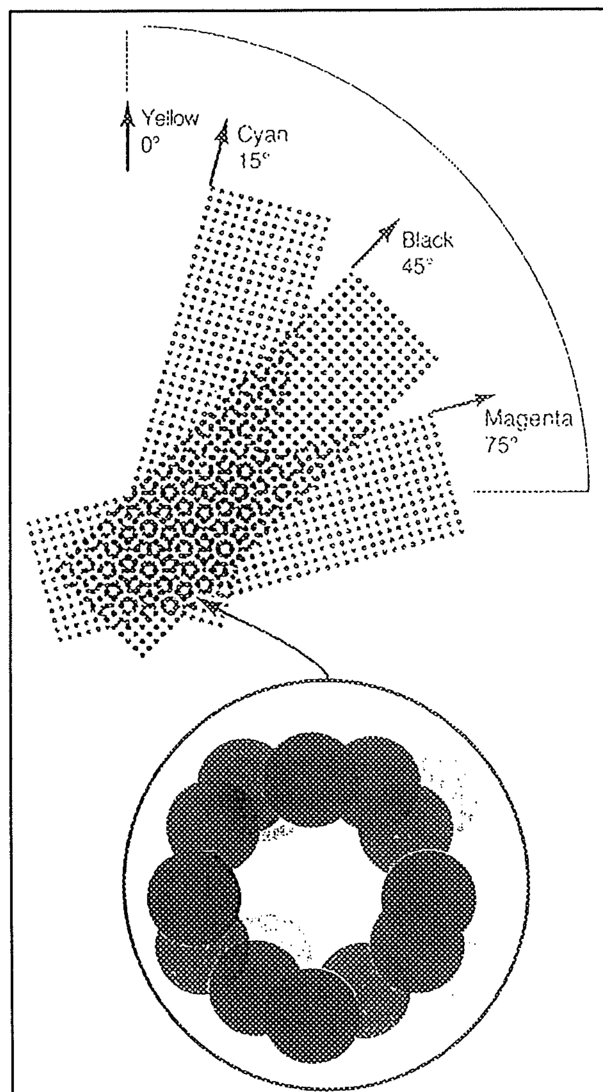
De optimale rasterhoeken
In de praktijk werkt men meestal met het rasterschema dat voorgesteld is in figuur 3/19.11-51. Eerst wordt het gele (Yellow) raster geprint, loodrecht op het papier. Nadien wordt het papier een tweede keer door de printer gevoerd en wordt het cyaan raster gedrukt. Dit raster heeft echter een rasterhoek van 15°. Nadien wordt het papier voor de derde keer bedrukt en wel met het maganta raster.
Ten opzichte van het gele raster heeft dit een hoek van 75°. Tot slot wordt het zwarte raster aangebracht en wel onder een hoek van 45°. Door de exacte keuze van de rasterhoeken ontstaan zogenoemde "rozetjes". Zo'n rozetje is in figuur

19.11 Principes en werking van printers

3/19.11-51 vergroot voorgesteld en is de basiscel, waarin de mengkleur ontstaat.



Figuur 3/19.11-50: Het verschijnsel "Moiré-effect" aan de hand van een voorbeeld toegelicht.



Figuur 3/19.11-51: De onderlinge positie van de vier rasters.

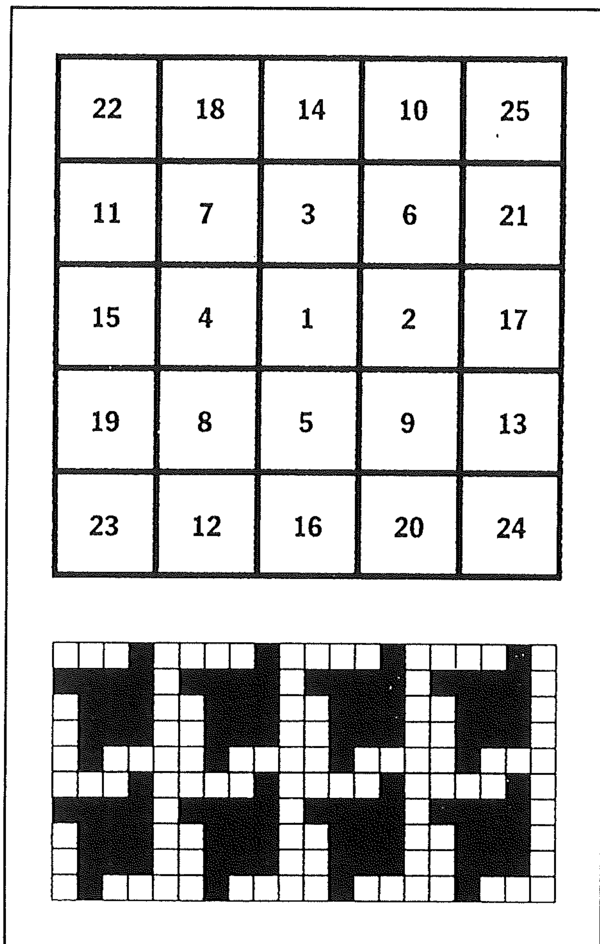
Rastercellen

Bij de meeste printers werkt men met rastercellen van vijf bij vijf inktdotjes. Dit wordt voorgesteld in figuur 3/19.11-52. Naarmate de intensiteit van de kleur moet toenemen, worden meer dotjes in de cel met inkt gevuld. Per kleur kan men dus 26 verschillende intensiteiten aanbrengen, van een volledig lege cel tot een volledig volle cel. Omdat men met drie kleurenrasters werkt kan men in totaal 26^3 is gelijk aan 17.576 verschillende mengkleuren samenstellen. Let op dat het zwarte raster niet meetelt bij het berekenen van het aantal af te drukken kleuren! Het zwarte raster wordt alleen gebruikt om de echt zwarte tinten in het ontwerp te printen.

De spot-functie

Een belangrijk begrip bij amplitude gemoduleerd rasteren in de zogenoemde spot-functie. Deze functie beschrijft hoe de 25 puntjes van een rastercel worden ingevuld naarmate de intensiteit van de kleur moet toenemen. In figuur 3/19.11-52 zijn de puntjes van de cel genummerd. Deze nummering geeft de volgorde aan waarmee de puntjes van de rastercel worden ingevuld bij toenemende intensiteit.

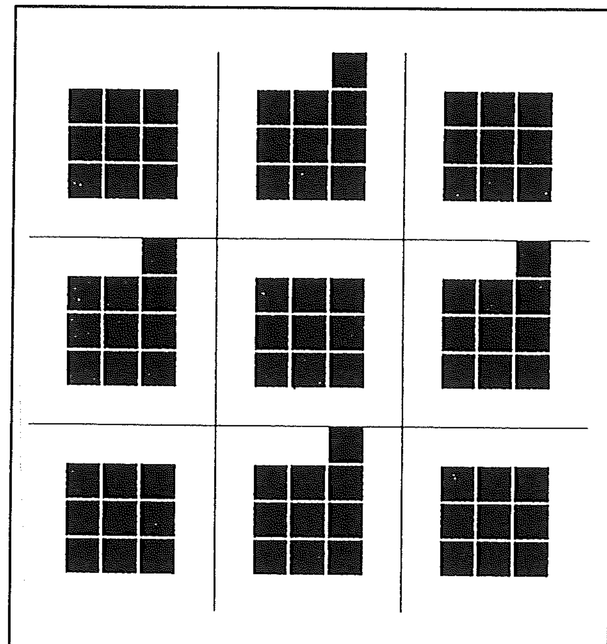
19.11 Principes en werking van printers



Figuur 3/19.11-52: De vijf bij vijf rastercel die bij AM-rasteren per inktkleur wordt toegepast.

Als voorbeeld wordt de invulling van de rastercel getekend als 12 van de 25 puntjes worden gevuld. Wie de volgnummers van de puntjes goed bestudeerd, zal vaststellen dat de rastercel min om meer spiraalvormig wordt ingevuld.

Sommige programma's, zoals "Illustrator" van Adobe, bieden de mogelijkheid de spot-functie van het raster aan te passen. Men kan dan voor ieder van de 25 mogelijke kleurintensiteiten een eigen indeling van al dan niet te printen inktdotjes definiëren. Ook in een printertaal zoals Postscript bestaat de mogelijkheid de spotfunctie van het raster vrij te definiëren.



Figuur 3/19.11-53: Het door Tektronix ontwikkelde principe van "adaptieve halftoning".

Overigens is het niet zo dat de spiraalvormige standaard spot-functie willekeurig gekozen is! In een jarenlange praktijk is gebleken dat deze de beste resultaten geeft.

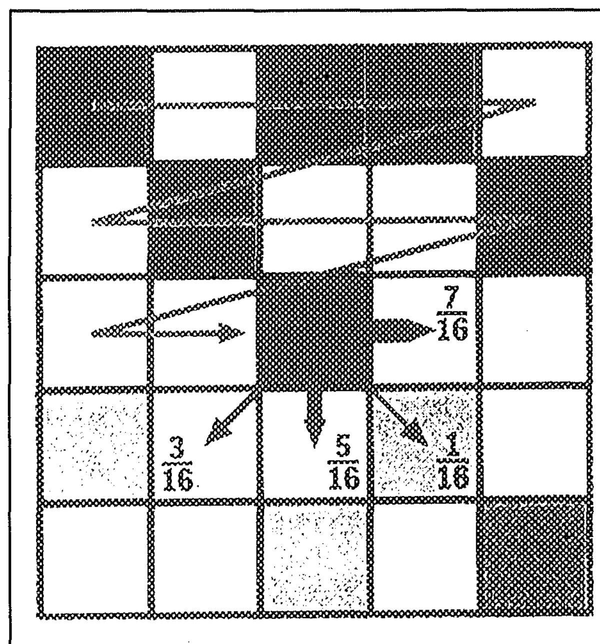
Adaptieve halftoning

Zoals beschreven kan een inktspuiter, die met AM-rastering werkt en met rastercellen van vijf bij vijf inktdotjes, in totaal 17.576 mengkleuren op het papier zetten. Dat aantal is uiteraard uit te breiden door de afmetingen van de rastercellen te vergroten. Maar het zal duidelijk zijn dat hierdoor de resolutie van de afdruk weer gaat dalen. Hoe groter de rastercellen, hoe minder details van het ontwerp op papier komen. Met een rastercel van vijf bij vijf kan men bij een 300 dpi printer een maximale rasterfrequentie van slechts ongeveer 60 lpi op het papier zetten. Als algemene regel geldt dus: hoe meer kleuren op papier, hoe lager de rasterfrequen-

tie, hoe hoger de rasterfrequentie hoe minder kleuren op papier.
Tektronix heeft in zijn 300 dpi kleuren laserprinters dit probleem op elegante manier opgelost. Het toegepaste procédé heet "adaptieve halftoning". Met deze techniek kan men het theoretisch bij een bepaalde rasterfrequentie horende maximaal aantal kleuren behoorlijk opvoeren. Tektronix werkt met rastercellen die ook vijf bij vijf dotjes groot zijn. Deze cellen worden echter verenigd tot een supercel van 15 bij 15 dotjes. Een en ander wordt voorgesteld in figuur 3/19.11-53. Bij het berekenen van de noodzakelijke invulling van de cellen wordt nu niet naar de afzonderlijke cellen gekeken, maar naar de supercellen. In figuur 3/19.11-53 wordt als voorbeeld de invulling van de supercel getekend als de software berekent dat 9,44 dotjes van de rastercellen met inkt gevuld moeten worden om de gewenste mengkleur af te drukken. Bij een normaal AM-raster zou de software dit moeten afronden naar 9 of naar 10 ingevulde dotjes. Door nu bepaalde dotjes van de rastercellen in de supercel wél en andere niet te vullen, kan men veel meer intensiteiten nabootsen. In feite zou men deze techniek een vorm van "printer enhancement technology" kunnen noemen, maar dan toegepast op kleur.

Frequentie gemoduleerd rasteren

Het grote probleem bij amplitude gemoduleerd rasteren is het vermijden van Moiré. Het regelmatige geometrische patroon van de rastercellen van de ene kleur interfereert met de regelmatige geometrische patronen van de rasters van de andere kleuren, waardoor ongewenste effecten kunnen ontstaan. Om dit probleem op te lossen heeft men gezocht naar een systeem, waarbij er geen regelmatige patronen in de rasters ontstaan. Men kan dan de vier rasters onder willekeurige hoeken over elkaar drukken, zonder kans te lopen dat Moiré-patronen ontstaan. Deze rastersystemen worden samengevat onder de naam "frequentie gemoduleerd rasteren" of "stochastisch rasteren". Bij deze systemen werkt men niet met rastercellen, maar met de individuele inktdotjes van de printer. Via ingewikkelde wiskundige systemen wordt bepaald welke dot's wél en welke dot's niet geprint moeten worden om een bepaalde intensiteit van de kleur tot stand te brengen.

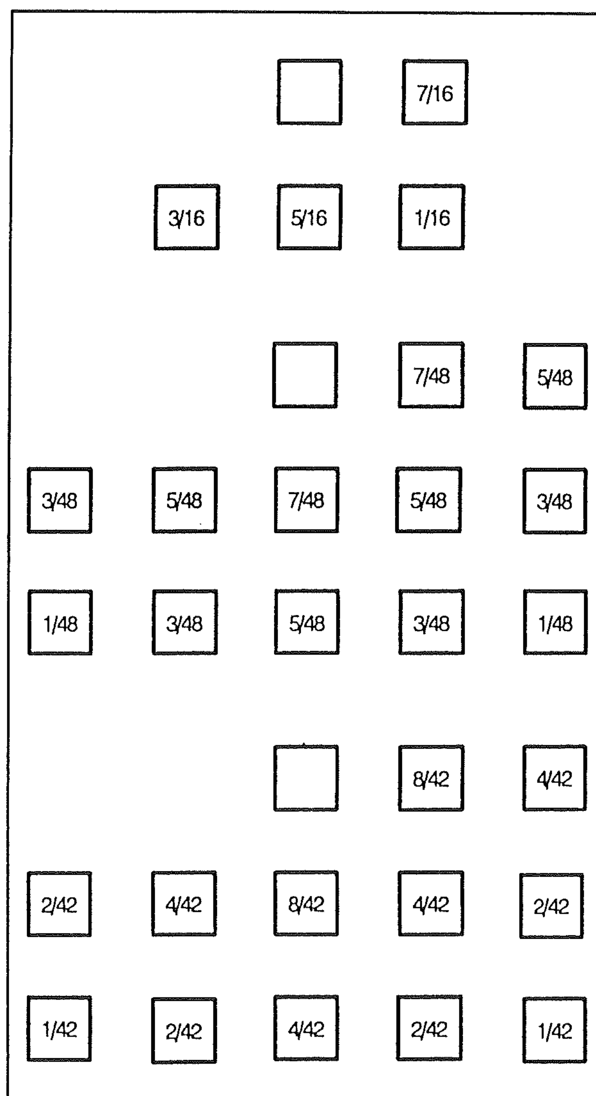


Figuur 3/19.11-54: Het principe van "error diffusion", de basis van frequentie gemoduleerd rasteren.

Error diffusion

Die ingewikkelde wiskundige algoritmen werken volgens het principe van de "error diffusion". Dit begrip, in het Nederlands te vertalen door "fout-diffusie" werd reeds kort besproken bij de dithering-technieken die bij zwart/wit foto's kunnen worden toegepast.

Bij het FM-rasteren van kleurenfoto's spelen deze technieken een zeer belangrijke rol. Reden om iets dieper op de werking in te gaan.
Bij het rasteren ontstaan steeds kleurfouten, omdat het in principe oneindig aantal kleurtinten moet omgezet worden in een beperkt praktisch haalbaar aantal. Bij AM-rasteren ontstaan deze afrondingsfouten natuurlijk in de rastercel zélf. Bij FM-rasteren wordt gepoogd deze fouten uit te smeren over een groter deel van de afdruk. Aan de hand van figuur 3/19.11-54 wordt een poging gewaagd de techniek van error diffusion toe te lichten.
In deze figuur zijn 25 naast elkaar liggende inktdotjes voorgesteld. Deze kunnen of geprint worden, of leeg blijven. De software behandelt de dotjes lijn na lijn, van links naar rechts. Als voorbeeld wordt het middelste dot besproken. De error diffusion techniek berekent, aan de hand van het origineel, de intensiteit van de te drukken kleur. Stel, dat de software berekent dat de intensiteit van de kleur in dit specifieke dot gelijk is aan 83 %. Omdat deze waarde boven 50 % ligt, wordt het dotje geprint. Zou de berekende intensiteit onder 50 % liggen, dan zou het dotje niet geprint worden. Er is nu echter een intensiteitsfout van niet minder dan 33 % opgetreden. Deze fout wordt verwerkt in de vier dot's die nog niet onder handen zijn genomen en die naast het te berekenen dot liggen. Deze vier dotjes zijn voorgesteld door middel van pijltjes. In ieder dotje is ook ingetekend met welke factor de fout in ieder dotje gecorrigeerd wordt. In het dotje dat er rechts naast ligt wordt de fout dus met 7/16 gecorrigeerd. Dat betekent dat de software, bij het berekenen van de invulling van dit dotje, 7/16 van de foutwaarde van het vorige dotje in de berekening verwerkt. Als zou blijken dat de intensiteit van dit dotje 45 % is, dan wordt hierbij 33 % * 7/16 = 14,43 % opgeteld, zijnde de correctiewaarde van het vorige dotje. In totaal wordt dus de gecorrigeerde intensiteit van dit dotje 59,43 %. Zonder error diffusiòn zou dit dotje dus niet geprint worden (intensiteitswaarde onder 50 %), met error diffusion wordt het wél geprint.



Figuur 3/19.11-55: De filter-matrixen van, van boven naar onder, Floyd-Steinberg, Jarvis-Judice-Ninke en Stucki.

19.11 Principes en werking van printers

Het begrip filter-matrix

Wat voor AM-rasters de spot-functie is, is voor FM-rasters de "filter-matrix". Die matrix definieert het aantal inktdotjes waarin wordt gecorrigeerd en de correctiefactoren die in die dotjes worden toegepast. Het zal duidelijk zijn dat de kwaliteit van de uiteindelijke kleurenafdruk via vier FM-rasters voor een belangrijk deel afhankelijk is van de toegepaste filter-matrix. Het besproken systeem, met vier correctiedotjes en coëfficiënten van 7/16, 5/16, 3/16 en 1/16, werd ontwikkeld door Floyd en Steinberg. Daarnaast zijn er tal van andere matrixen ontwikkeld en uitgeprobeerd, zoals deze van Jarvis-Judice-Ninke en deze van Stucki. De samenstellingen van deze drie bekendste filtermatrixen zijn voorgesteld in figuur 3/19.11-55.

Postscript Level 2

Tot begin 1994 gebruikte iedere printerfabrikant en softwareleverancier, die zich bezig hield met het implementeren van FM-rasters in zijn producten, wel een eigen filter-matrix.
Enige bekende namen:
- FinePoint van Tektronix, reeds in 1991 in sommige 300 dpi kleurenlasers van deze firma ingebouwd;
- Cristal Raster van Agfa;
- Diamond Screening van Linotype;
- Lazel van Crosfield.

Een ongewenste situatie! Eind 1994 kwam Adobe echter met een nieuwe versie van Postscript Level 2 op de markt, waarin ook de FM-technologie werd ondersteund. Het toegepaste wiskundige algoritme werd "Adobe Brilliant Screening" genoemd. Vanwege de hoge marktpenetratie van Postscript maakt deze FM-techniek grote kans om een internationale standaard te worden.

Voordelen van FM-rasteren

FM-rasters hebben tal van voordelen, waarvan de voornaamste nu in het kort worden samengevat.
- Geen interferenties
  Omdat gedrukt wordt zonder geometrische rasters, ontstaan geen interferentie-patronen, zoals Moiré. Ook interferentie met het beeld zélf, zoals de afdruk van een kostuum met een visgraat motief, is onmogelijk.
- Geen zichtbare rozetten
  De bij AM-rasters zo specifiek aanwezige rozetjes (zie figuur 3/19.11-51) ontbreken volledig.
- Meer details
  Bij een bepaalde rasterfrequentie lijkt het alsof FM-rasters meer details weergeven dan AM-rasters. Blijkbaar wordt het menselijk oog minder afgeleid door de bij FM afwezige rastergeometrie en is daardoor in staat een hogere scherpte-indruk te vormen.
- Zuiverder kleurverlopen
  Kleurverlopen, afgedrukt met FM-rasters, zien er gladder uit dan bij AM-rasters.
- Minder registerfouten
  Bij AM-drukwerk valt het onmiddellijk op als een van de rasters niet nauwkeurig in de juiste positie is geprint. Dit noemt men "registerfouten". Die fouten uiten zich in storende beeldranden of onscherpte. Dezelfde registerfouten bij FM zijn veel minder zichtbaar.

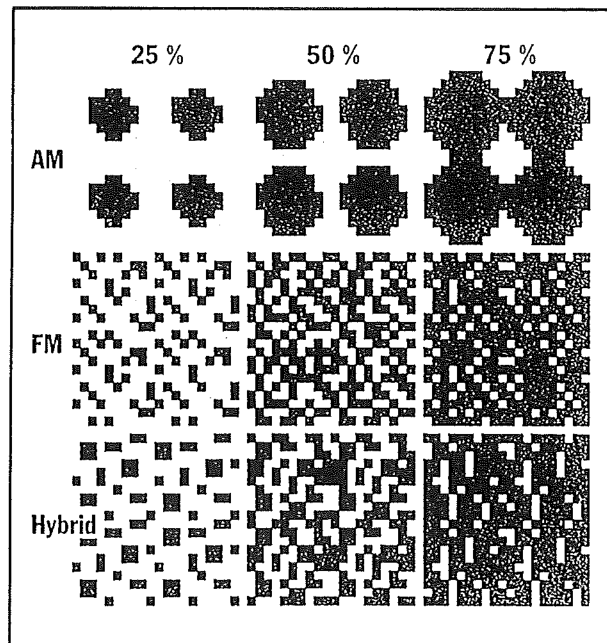
Nadelen van FM

Frequentie gemoduleerd rasteren heeft echter ook nadelen.
- Korreligheid
  Kleurvlakken in een egale kleur of met weinig kleuren vertonen vaak een hinderlijke korreligheid, een verschijnsel dat bij AM niet aanwezig is.

– Puntverbreding
  Bij het afdrukken op een printer zullen de inktdotjes altijd iets groter zijn dan eigenlijk zou moeten. Dit wordt veroorzaakt door het uitsmeren van de kleuren over het papier of door het platwalsen van de tonerdotjes bij laserprinters. Een verschijnsel dat "dot gain" wordt genoemd. Dit verschijnsel speelt ook bij AM een belangrijke rol, maar blijkt bij FM nog veel kritischer te zijn. In het algemeen blijkt dat bij FM-rasters een twee maal grotere puntverbreding optreedt dan bij AM-rasters. Hetgeen betekent dat foto's veel donkerder worden afgedrukt dan men gewend is en men dus nog sterker helderheid en contrast moet corrigeren dan bij AM gebruikelijk is.
– Meer tonerverbruik
  Uit praktijkproeven blijkt dat FM-afdrukken veel meer toner of inkt gebruiken dan AM-afdrukken. De prijs per afdruk ligt dus iets hoger.

Hybride rasters
Tot slot nog enige zinnen over hybride rasters. Bij deze techniek worden combinaties berekend van traditionele AM-rasters en moderne FM-rasters. Een hybride raster kenmerkt zich zowel door variabele afstand tussen de geprinte dotjes (FM-typisch) als door variabele afmetingen van de geprinte inktvlekken (AM-typisch). Op deze manier probeert men de voordelen van beide systemen met elkaar te combineren.

Vergelijking
Tot slot van de bespreking van rastertechnieken voor kleurenprinters wordt in figuur 3/19.11-56 een vergelijking geschetst tussen een AM-raster, een FM-raster en een hybride raster voor kleurintensiteiten van 25 %, 50 % en 75 %. Een zeer heldere samenvatting, waarbij de principiële verschillen tussen de drie systemen onmiddellijk in het oog springen!



Figuur 3/19.11-56: Het principiële verschil tussen AM-, FM- en hybride rasters kort en krachtig samengevat.

Van beeldscherm naar printer

Beeldscherm/printer compatibiliteit
In een vorige paragraaf is al één reden aangevoerd voor het probleem van de beeldscherm/printer incompatibiliteit: de Pantone-codering. Daarnaast is er echter nog een aantal factoren, dat er toe bijdraagt dat de resultaten op papier afwijken van de resultaten op het beeldscherm:
– de ijking van het beeldscherm;

19.11 Principes en werking van printers

- de eigenschappen van het papier;
- de kleurwaarneming.

Deze factoren worden in de volgende paragraafjes toegelicht.

De ijking van het beeldscherm

Het probleem met een beeldscherm is dat een dergelijk apparaat twee knoppen heeft, waarmee de intensiteit en het contrast kunnen worden ingesteld. Bovendien heeft een beeldscherm een bepaalde verzadigingsgrens. Draait men de intensiteit helemaal open, dan worden er zoveel elektronen naar de fosforpuntjes gestuurd, dat deze hun maximale lichtintensiteit uitstralen, zelfs bij minder heldere kleuren. Heel helder rood en iets minder helder rood verschijnen dan als dezelfde tint op het scherm. Het gevolg is dat men de kleuren op het scherm zo donker of zo licht kan instellen als men zelf wil. Een in principe licht gele kleur kan dus, door een verkeerde instelling van contrast en intensiteit, donker geel op het beeldscherm verschijnen.

De printersoftware weet echter niets van deze monitor instellingen. Deze software houdt alleen rekening met de kleurcodering die het in de file aantreft. Dit verschijnsel is een van de voornaamste redenen waarom een afdruk op papier zo kan afwijken van de kleuren die men op het beeldscherm ziet. De oplossing voor dit probleem is gebruik te maken van beeldijking. Goede grafische software heeft een optie ingebouwd om de kleuren op het scherm te ijken. In de meest eenvoudige vorm wordt een testfile meegeleverd, die een kleurenkaart bevat. Die kleurenkaart bestaat bijvoorbeeld uit 16 blokjes rood, ieder met een andere intensiteit. Laadt men deze file, dan is het de bedoeling dat men deze 16 blokjes duidelijk van elkaar kan onderscheiden. Is dat niet het geval, dan moet men contrast en intensiteit van het beeldscherm bijregelen, tot de zestien blokjes duidelijk ieder een andere kleurintensiteit hebben.

De eigenschappen van het papier

De eigenschappen van het gebruikte papier bepalen in grote mate de manier waarop een kleurontwerp wordt afgedrukt. Het papier is immers de “indirecte lichtbron”, die door min of meer te reflecteren de uiteindelijke kleurindruk bepaalt. Hoe gladder en glanzender het oppervlak van het papier is, hoe helderder en stralender de kleurtinten lijken. Als men hetzelfde ontwerp op ruwer en grijzer papier afdrukt, zullen de stralende kleuren veel fletser lijken. Dat is logisch, want het ruwe en grijze papieroppervlak kaatst minder invallend licht terug. Bovendien zal dit papier de golflengte samenstelling van het teruggekaatste licht veranderen, waardoor er een kleurverschuiving optreedt.

Vandaar dat kleurontwerpen steeds moeten worden afgedrukt op papier dat “effen wit” wordt genoemd. Dit papier heeft de eigenschap dat de golflengte samenstelling van het teruggekaatste licht precies gelijk is aan deze van het invallend licht.

Een tweede probleem ontstaat door de oppervlaktestructuur van het papier. Als het papier veel houtvezels bevat ontstaat een verschijnsel dat het “spinneweb-effect” wordt genoemd. De inkt zal zich dan langs deze vezels verspreiden, waardoor de in principe ronde inktspatjes zich uitbreiden tot een soort van miniatuur spinneweb. Het gevolg is echter dat naast elkaar liggende inktspatjes zich gaan vermengen, waardoor plotselinge kleurovergangen op het scherm in elkaar overvloeien op het papier. Dit noemt men het “mixen” van kleuren.

Besluitend kan men stellen, dat de kwaliteit van het papier een zeer grote rol speelt in de uiteindelijke kleurindruk van een afdruk.
Aanbevolen wordt steeds gebruik te maken van de speciale papiersoorten, die door de printerfabrikant worden geleverd dan wel worden aanbevolen, ook al zijn deze meestal zeer prijzig.

De kleurwaarneming
Kleuren zijn niet statisch, maar dynamisch. De visuele waarneming van een kleurtint hangt van een heleboel externe factoren af, zoals:

- De verlichting
  Bij een beeldscherm spelen deze factoren een minder belangrijke rol, omdat het beeldscherm een lichtbron is. Het papier is echter alleen een weerkaatser van het licht en die weerkaatsing is uiteraard afhankelijk van het soort licht dat wordt gebruikt. Een kleurafdruk op papier zal er onder daglicht heel anders uitzien dan onder het licht van gloeilampen. Het licht van de zon bevat veel blauwachtige golflengten, terwijl het licht van gloeilampen veel meer rode golflengten heeft. Het papier zal dus in het ene geval veel blauwe golflengten terugkaatsen en in het andere geval veel rode.
  De totaalindruk van het beeld zal in het ene geval een blauw- en in het andere geval een roodverschuiving bevatten. In vaktermen wordt dit verschijnsel "metametrische kleurverschuiving" genoemd.
- De achtergrond
  Ook de achtergrond, waartegen met een kleurontwerp bekijkt, speelt een belangrijke rol bij de kleurindruk. Hiervoor bestaat een aantal eenvoudige regels:
  - donkere kleuren lijken donkerder tegen een donkere achtergrond dan tegen een lichte achtergrond;
  - complementaire kleuren kunnen het best niet gemengd worden, bijvoorbeeld rode kleuren op groen papier, want het resultaat lijkt vaal;
  - wanneer complementaire kleuren naast elkaar worden toegepast, bijvoorbeeld een groen affiche op een rode muur, versterken zij elkaars effect waardoor het resultaat vaak pijn doet aan de ogen.
- Het papier
  De invloed van het papier op de kleurwaarneming werd reeds eerder beschreven. Er is echter toch nog een punt dat de aandacht verdient. Tegenwoordig kan men papier bestellen in ongeveer alle kleuren van de regenboog. Zwart/wit ontwerpen komen op bijna iedere kleur papier tot hun recht. Met kleurontwerpen ligt dat echter duidelijk anders. Een stelregel: gebruik steeds wit papier, omdat de invloed van gekleurd papier op een gekleurd ontwerp erg moeilijk is in te schatten!

Printertalen

Inleiding
Printers worden vanuit de PC aangestuurd door middel van stuurprogramma's. Deze stuurprogramma's zetten de te printen informatie om in signalen die de printer verstaat. In de loop der tientallen jaren zijn er ontelbare printertalen op de markt verschenen, waarvan de meesten weer geruisloos zijn verdwenen. Een aantal zijn echter de-facto standaarden of echte standaarden geworden, zoals:
- ESC/P;

19.11 Principes en werking van printers

- HPGL;
- PCL;
- PreScribe;
- Postscript;
- TrueImage;
- GDI.

Soorten
Printertalen kunnen ingedeeld worden in twee categorieën:
- PDL's;
- non-PDL's.

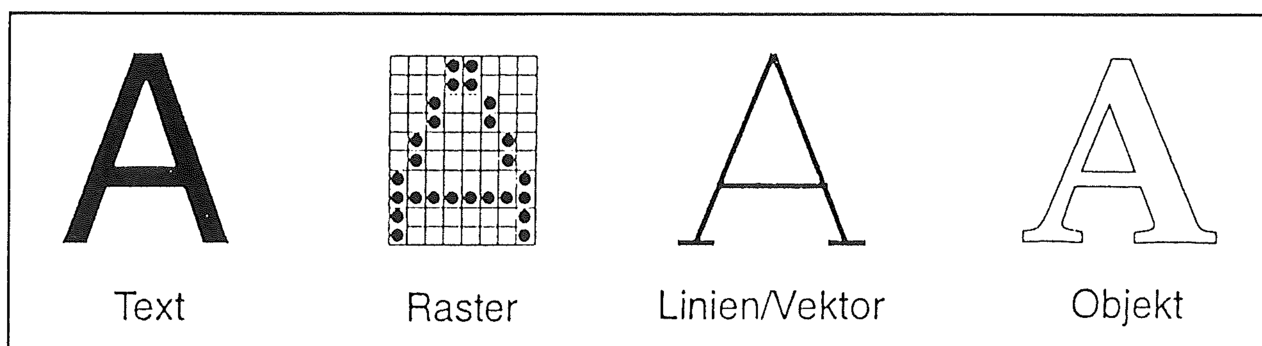
PDL is de afkorting van "Page Description Language". Printertalen die volgens dit principe werken, sturen hun printer altijd pagina-georiënteerd aan. Het zal duidelijk zijn dat men dergelijke talen alleen kan gebruiken bij pagina-georiënteerde printers, zoals laserprinters. De meeste PDL's zijn interpretertalen. Dat betekent dat de commando's in min of meer begrijpelijk Engels als zuivere tekstfile worden geschreven en als dusdanig naar de printer gestuurd. In de hardware van de printer worden deze commando's geïnterpreteerd en op machinecode niveau omgezet in instructies voor de speciale processor, de RIP (Raster Image Processor) die verantwoordelijk is voor het dot na dot samenstellen van het volledige beeld van een pagina in het geheugen van de printer. In feite zijn dergelijke talen volledig vergelijkbaar met BASIC of PASCAL, met als enig verschil dat de commando-set volledig is toegespitst op het geven van grafische instructies. Non-PDL talen werken ook met instructies, die echter niet onder de vorm van een tekstfile naar de printer gestuurd kunnen worden, maar onder de vorm van speciale binaire code's. De hardware in de printer ontdekt deze speciale code's aan hun samenstelling en voert een eenvoudige actie uit, zoals het omschakelen van normaal schrift naar smal schrift. Het zal duidelijk zijn dat deze talen voornamelijk worden toegepast bij regelgeoriënteerde printers.

Printertalen en letters
Nog steeds is het zo dat het meeste dat geprint wordt bestaat uit tekst, dus uit letters. Vandaar dat de verwerking van letters door een printertaal een zeer belangrijk gegeven is. Hiervoor heeft men drie verschillende systemen ontwikkeld:
- raster-georiënteerde letters;
- lijnen-georiënteerde letters;
- object-georiënteerde letters.

De verschillen tussen deze drie systemen worden kernachtig toegelicht aan de hand van figuur 3/19.11-57. Als voorbeeld wordt de letter "A" behandeld.

Raster-georiënteerd
Bij de rastertechnologie wordt er als het ware een lege matrix, bestaande uit x bij y cellen, over de letter gelegd en bekeken welke cellen gevuld moeten worden om de letter zo goed mogelijk te benaderen. Deze informatie wordt opgeslagen in het geheugen van de printer en wordt uitgelezen als de printertaal de instructie geeft om de letter op papier te zetten. Gevulde rastercellen worden vertaald naar te printen inktdot's. Het voordeel is dat gerasterde letters zeer snel geprint kunnen worden, omdat de printer alleen gegevens vanuit zijn geheugen naar het papier moet overbrengen. Er zijn geen ingewikkelde berekeningen noodzakelijk. Maar anderzijds heeft deze technologie veel nadelen. Op de eerste plaats kost het bewaren van alle informatie van een volledig font verschrikkelijk veel geheugenruimte. Op de tweede plaats kunnen zij weliswaar vergroot of verkleind worden afgedrukt, maar dat gaat nooit zonder verlies van informatie.

19.11 Principes en werking van printers



Figuur 3/19.11-57: Printertalen kunnen letters op drie manieren weergeven. Van links naar rechts: als raster, als lijnen of als object.

De printertaal kan weliswaar uit de opgeslagen rasterinformatie de printer zo aansturen dat de letters op dubbele grootte worden afgedrukt, maar dan worden de letters nog hoekiger dan zij al zijn.
Rasterletters worden voornamelijk toegepast in naald matrixprinters die via ESC/P worden aangestuurd, maar ook in laserprinters die werken met de allereerste versies van PCL.

Lijnen-georiënteerd

De letters worden op het papier gezet door het tekenen van lijntjes, cirkels en bogen. De definitie van de volledige letter kan dus opgeslagen worden onder de vorm van een aantal eenvoudige tekeninstructies. Via wiskunde is het mogelijk de letters niet alleen vergroot en verkleind te printen, maar zij kunnen bijvoorbeeld ook onder een bepaalde hoek op het papier worden gezet.
Het nadeel is dat de letters er vrij dunnetjes uitzien. Volle vlakken kunnen alleen gemaakt worden door een aantal lijntjes naast elkaar te tekenen. Lijnen-georiënteerde letters worden alleen toegepast in HPGL.

Object-georiënteerd

Bij deze techniek wordt een letter opgevat als een te tekenen object, dat is opgebouwd uit lege of volle cirkels, vierkantjes, rechthoeken, lijnen en bogen. Allereerst wordt de omtrek, de zogenoemde "outline" van de letter vastgelegd in een aantal wiskundige formules. Hiervoor worden niet alleen rechte lijnen en cirkelsegmenten gebruikt, maar ook "Bézier"-curves of "B-Spline"-curves, niet regelmatig verlopende lijnen die toch wiskundig volledig te definiëren zijn. Nadien worden wiskundige algoritmen toegepast om de outline van de letter zo snel mogelijk te vullen. De letter wordt nu volledig beschreven door een stelsel van wiskundige vergelijkingen. De printertaal zorgt er nu voor dat deze wiskundige vergelijkingen met de hoogst mogelijke resolutie worden afgedrukt op het papier. Kortom: een object-georiënteerde letter wordt op een 300 dpi printer afgedrukt met een resolutie van 300 dpi, maar op een 1.000 dpi printer met een resolutie van 1.000 dpi. Het zal duidelijk zijn dat object-georiënteerde letters het fraaist op papier komen, maar dat er heel wat rekenwerk bij komt kijken om de in het geheugen van de printer aanwezige formules om te rekenen naar al dan niet te printen inktdot's. Deze lettertechnologie wordt alleen ondersteund door hoog ontwikkelde printertalen, zoals Postscript, PreScribe en de allernieuwste versies van PCL.

19.11 Principes en werking van printers

```
10 REM BASIC program example
20 LPRINT CHR$(27)+"&l8DYou can change line spacing."
30 LPRINT "Notice that these three lines are"
40 LPRINT "printed at 8 lines per inch."
50 LPRINT
60 LPRINT CHR$(27)+"&l6DYou can "+CHR$(27)+"(s1B";
70 LPRINT "turn on boldface"+CHR$(27)+"(s0B"
80 LPRINT CHR$(27)+"&dDor underline words."+CHR$(27)+"&d@"
90 LPRINT "You can even "+CHR$(27)+"&k2S";
100 LPRINT "change the print pitch."+CHR$(27)+"&k0S"
110 END
```

Output

```
You can change line spacing.
Notice that these three lines are
printed at 8 lines per inch.

You can turn on boldface
or underline words.
You can even change the print pitch.
```

Figuur 3/19.11-58: Een voorbeeldje van het gebruik van Escape-sequenties in een BASIC-programma.

ESC/P

Deze reeds in 1981 door Epson gepatenteerde non-PDL taal maakt gebruik van eenvoudige ASCII-commando's die allemaal met het "Escape"-teken met ASCII-code "27" beginnen. Deze taal werd en wordt voornamelijk gebruikt voor het doorgeven van eenvoudige instructies naar naald matrixprinters. Een Escape-code kan bijvoorbeeld gebruikt worden om de printer om te schakelen van sneldruk naar schoondruk of om de breedte van de letters in te stellen. Dergelijke zogenoemde Escape-sequenties kunnen vanuit software, bijvoorbeeld een tekstverwerker, naar de printer gestuurd worden, kunnen geprogrammeerd worden in een simpel BASIC-programmaatje door middel van LPRINT-opdrachten of kunnen rechtstreeks vanaf het toetsenbord naar de printer gestuurd worden.

In figuur 3/19.11-58 is als voorbeeld een BASIC-programma weergegeven en de print-resultaten op een printer die volgens het ESC/P-protocol werkt.

Het oorspronkelijke zeer eenvoudige ESC/P-protocol werd door een heleboel printerfabrikanten geadopteerd en naar eigen behoeften en mogelijkheden van hun producten uitgebreid. Zo kan men bij diverse printers zelfs een volledig nieuw font (lettertype) met ESC-codes definiëren en opbergen in het geheugen van de printer. Bovendien kan men zelfs alle

19.11 Principes en werking van printers

naaldjes van de printer afzonderlijk aansturen, waardoor het mogelijk is een programma te schrijven voor het afdrukken van bijvoorbeeld een bedrijfslogo. Maar natuurlijk zijn al die mogelijkheden volledig achterhaald, nu zo ongeveer ieder programma in staat is de inhoud van het beeldscherm dot voor dot naar een printer te sturen.

HPGL

HPGL is de afkorting van "Hewlett Packard Graphic Language". Deze non-PDL taal werd oorspronkelijk ontwikkeld voor het aansturen van de grafische plotter HP7475A die door HP op de markt werd gebracht. Een plotter werkt met een aantal pennen, die van links naar rechts over het papier bewogen kunnen worden. Maar daarnaast kan het papier ook nog eens op en neer gestuurd worden, zodat de pen in staat is ieder vlakje van het papier te vullen met een lijn. Met een HPGL-instructie kan men bijvoorbeeld de plotter opdracht geven ergens op het papier een cirkel te plotten. De originele versie van HPGL is een non-PDL taal. Men kan bijvoorbeeld eerst een instructie geven om een cirkel te plotten en nadien een instructie om ergens op de pagina een lijn te tekenen. De plotter zal eerst de cirkel tekenen en nadien de lijn. Bij plotters kan dat, omdat deze apparaten het papier volledig vrij heen en weer kunnen laten bewegen onder het penmechanisme.
Diverse printerfabrikanten hebben deze taal echter ook geïmplementeerd in hun pagina-georiënteerde printers. Natuurlijk kunnen niet alle HPGL-commando's geïnterpreteerd worden en is het ook niet mogelijk commando na commando in te geven en de resultaten op papier te bewonderen. Bij een laserprinter wordt het papier immers maar één keer doorgevoerd en alle instructies moeten dus in deze ene doorgang verwerkt worden. Vandaar dat een extra commando werd ingevoerd, AF (Advance Frame), waarmee de printer te verstaan wordt gegeven dat alle HPGL-commando's kunnen geïnterpreteerd worden en afgedrukt op een vel papier. Commando's voor het optillen van de pen of voor het wisselen van de penkleur worden natuurlijk ook niet ondersteund.
De commando's van HPGL worden ingevoerd als tweeletterige codes, voorzien van parameters en afgesloten met puntkomma's.
Als voorbeeld wordt de instructie behandeld, waarmee men een boog kan tekenen. De syntax van deze instructie is:
AA x,y,angle,chord;
waarin:
- AA:
  de naam van de instructie, in dit geval "Arc Absolute";
- x,y:
  de absolute coördinaten van het middelpunt van de boog, waarbij bij A4-papier wordt uitgegaan van 603 bij 521 eenheden voor het gehele vel;
- angle:
  het aantal graden waaronder de boog getekend wordt;
- chord:
  een positief getal dat de tekennauwkeurigheid van de boog definieert.

PCL

PCL is de afkorting van "Printer Control Language". Ook deze taal werd door Hewlett Packard geïntroduceerd en wel in 1990 bij de introductie van de Laserjet III. Een van de belangrijkste innovaties van deze speciaal voor laserprinters ontwikkelde taal was de mogelijkheden letter-

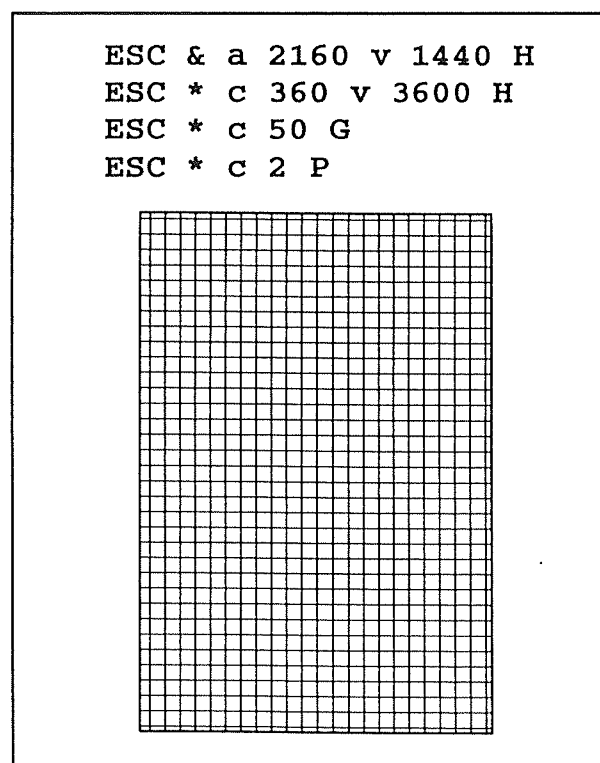
fonts op een eenvoudige manier te "downloaden". Men kon dus andere soorten letters gebruiken dan deze die standaard in het geheugen van de printer waren ingebakken.
In de allereerste versies van PCL werden alleen raster-georiënteerde letters ondersteund.
Ondertussen is Hewlett Packard echter reeds toe aan PCL-5. Deze versie is volledig downwards-compatibel met de vorige versies PCL-1 tot en met PCL-4, maar beschikt over een aantal extra's, zoals:

- ingebouwde interpreter voor HPGL-instructies;
- ondersteuning van object-georiënteerde fonts, die volledig schaalbaar en draaibaar zijn.

PCL-5 bezit lang niet alle mogelijkheden van Postscript, maar is een taal die alle noodzakelijke praktische mogelijkheden biedt die van een printertaal in een kantooromgeving worden verwacht.
Een specifieke eigenschap van PCL is de mogelijkheid macro's samen te stellen. Een macro bestaat uit een aantal samenhangende instructies, die in één keer naar de printer verstuurd kunnen worden. De macro's worden opgeslagen in het gebruikersgeheugen van de printer en kunnen dus steeds weer eenvoudig aangeroepen worden.
Daarnaast heeft PCL tamelijk uitgebreide vul-instructies.
Een gesloten vlak kan gevuld worden met patronen en met grijsrasters met bepaalde densiteiten.
De commando's van PCL bestaan, net zoals die van ESC/P, uit Escape-sequenties. Als voorbeeld wordt in figuur 3/19.11-59 de Escape-sequenties gegeven die de printer opdragen een vlak te printen, dat gevuld wordt met een bepaald patroon.

- regel 1:
  verplaats de "cursor" naar een startpositie;
- regel 2:
  definieer de afmetingen van de rechthoek;
- regel 3:
  definieer het vulpatroon;
- regel 4:
  print de rechthoek.

Uit dit voorbeeld blijkt duidelijk dat PCL nu niet bepaald een gemakkelijke taal is om mee te programmeren!



Figuur 3/19.11-59: Een voorbeeldje van het programmeren van een gevuld vlak met PCL-instructies.

PreScribe
PreScribe is een PDL taal die ontwikkeld werd door Kyocera. Het grote voordeel van deze taal is dat de instructies worden aangegeven door sleutelwoorden, die

19.11 Principes en werking van printers

min of meer duidelijk maken wat de bedoeling van een instructie is. Deze taal is dus min of meer vergelijkbaar met BASIC, met instructies als:
- BOX;
- CIR;
- TEXT;
- FONT;

die onmiddellijk duidelijk zijn. Deze sleutelwoorden moeten natuurlijk worden aangevuld met parameters en afgesloten met een punt-komma.
PreScribe ondersteunt 79 rastergeoriënteerde lettersoorten en 3 wat genoemd wordt "dynamische fonts". Deze drie fonts zijn object-georiënteerd en kunnen dus geschaald, geroteerd, gerasterd en geschaduwd worden. Daarnaast beheerst de taal alle noodzakelijke instructies voor het tekenen van lijnen, cirkels, blokken, etc. Een bijzonderheid van PreScribe is dat er instructies ter beschikking staan voor het printen van BAR-codes, waarbij niet minder dan 41 soorten ondersteund worden, waaronder de in Europa gestandaardiseerde EAN. In 1992 werd PreScribe II gelanceerd. Deze versie ondersteunt 13 objectgeoriënteerde fonts en de mogelijkheid logo's, handtekeningen, etc. onder een macro onder te brengen. Door het aanroepen van zo'n macro wordt een handtekening op een gedefinieerde plaats van de pagina geprint. Printers die met PreScribe II werken, bevatten een gleuf voor het aanbrengen van "IC-card's". Dit zijn op creditcard's lijkende kaartjes, waarin 64 kB of 128 kB geheugen aanwezig is. Deze kaartjes zijn bedoeld voor het opslaan van deze macro's en de inhoud van de kaartje kan met enige eenvoudige instructies naar het papier verzonden worden. Verder zijn er een aantal instructies bijgevoegd, waarmee men op een heel eenvoudige manier formulieren kan vormgeven. De instructies van PreScribe kunnen als ASCII-tekstfile met een tekstverwerker geschreven worden en worden door de printer als dusdanig opgevat door de tekencombinatie >>!R!<< naar de printer te zenden. Dat is een combinatie die wel nooit in teksten zal voorkomen en dus door de taal eenduidig als het begin van een instructietekst kan worden opgevat. Na het verzenden van de ASCII-file wordt de printer weer naar normale modus geschakeld door het verzenden van de in teksten al even ongebruikelijke tekencombinatie >>exit;<<. Het grote voordeel van deze gang van zaken is dat men de printer vanuit de tekstverwerker kan programmeren! Het enige waarop men moet letten is dat de tekstverwerker zuivere ASCII-files naar de parallelle poort stuurt. In figuur 3/19.11-60 is een voorbeeldje van een Prescribe-programma gegeven, met de printresultaten.

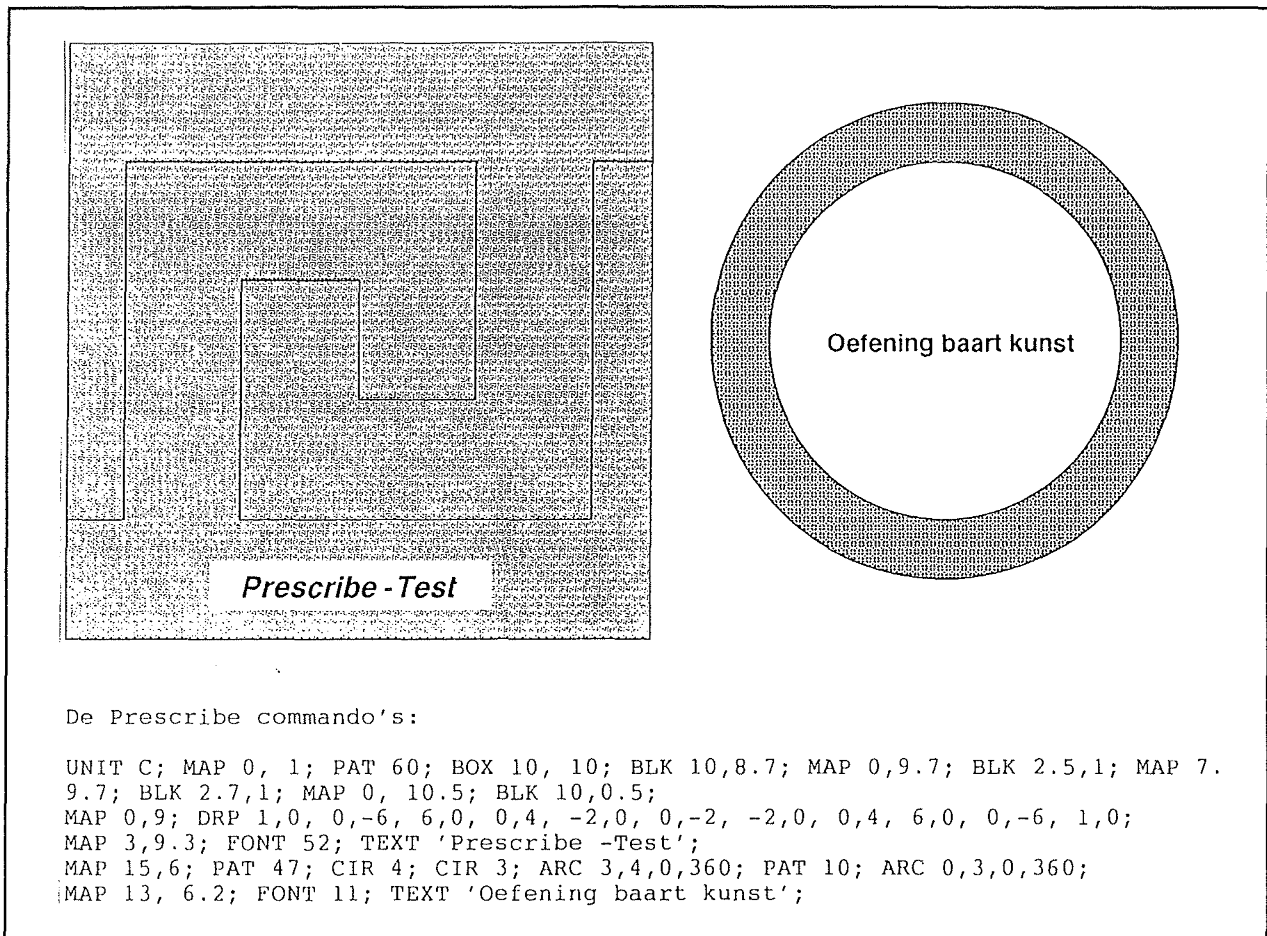
Postscript

Postscript is een paginabeschrijvingstaal, die door Adobe werd ontwikkeld en uitgegroeid is tot dé standaard op het gebied van printertalen. Alle bekende printerfabrikanten hebben laserprinters in de handel, die Postscript-gestuurd moeten worden.
Alle goede tekst- en opmaakprogramma's ondersteunen de mogelijkheid om pagina's als PS-file naar de printer te sturen. Postscript is een dynamische PDL. Dat betekent dat de taal toelaat controlestructuren als:
- IF . . . THEN;
- FOR . . . NEXT;
- REPEAT;

in de paginabeschrijving op te nemen, waardoor het programmeren van ingewikkelde printstructuren heel erg vereenvoudigd kan worden.

wordt vervolgd

19.11 Principes en werking van printers



Figuur 3/19.11-60: Een programmavoorbeeldje, geschreven in PreScribe.

Natuurlijk ondersteunt Postscript object-georiënteerde fonts, waarvan er ondertussen vele duizenden als Type-1 op de markt zijn.
De allereerste versie, Level-1, werd al snel uitgebreid met drie losse modulen.

- De Color Extensions
  Hiermee kan men via Postscript rechtstreeks kleurenprinters volgens het CYMK-model aansturen. De omrekening RGB/CYMK moest namelijk vóór het beschikbaar komen van deze uitbreiding door externe software gebeuren.
- Display Postscript
  Met deze uitbreiding wordt ook de opbouw van een pagina op het scherm door Postscript aangestuurd. Het gevolg hiervan is dat Postscript absolute WYSIWYG kan garanderen. Wat op het scherm staat verschijnt op precies dezelfde manier op het papier! Een hele verbetering ten opzichte van andere systemen, waar de vertaling van de gegevens op het scherm, via een printertaal, naar de gegevens op papier vaak erg gebrekkig verloopt. Wél moet opgemerkt worden, dat Display Postscript erg veel rekencapaciteit van de processor eist. Vandaar dat deze extensie niet erg populair is geworden in de PC-wereld, die immers jarenlang geteisterd werd door veel te trage Intel-processoren.

19.11 Principes en werking van printers

- Composite Fonts
  Via deze uitbreiding kunnen zeer ingewikkelde fonts, die uit heel veel karakters bestaan, bijvoorbeeld Chinees of Japans, toch op een zeer economische manier geprogrammeerd worden.

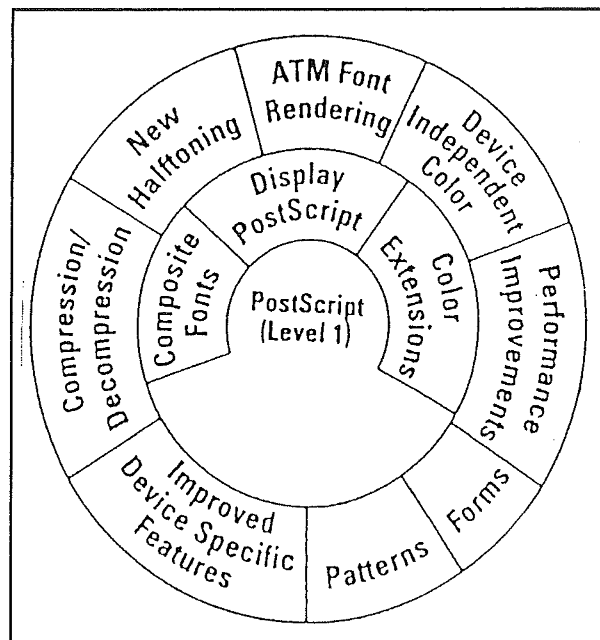
In mei 1991 werd door de ISO, de "International Standards Organisation" een internationaal gestandaardiseerde paginabeschrijvingstaal ingevoerd: de SPDL, oftewel "Standarized Page Description Language". Deze taal werd ontwikkeld met de hulp van specialist Adobe en het zal dan wel geen verbazing wekken dat deze internationale standaard grotendeels is gebaseerd op Postscript Level-1. Hiermee werd Postscript van een de-facto standaard een officieel erkende standaard printertaal.

In 1991 werd Level-2 van Postscript door Adobe geïntroduceerd. Dit is een volledig met Level-1 downwards-compatibele uitgebreide versie van Postscript. De drie reeds beschreven extensions werden hierin geïntegreerd, maar daarnaast werd Level-2 nog eens uitgebreid met niet minder dan acht nieuwe uitbreidingen.

Alle features van Postscript Level-2 zijn overzichtelijk samengevat in figuur 3/19.11-61.

Een korte bespreking van de voornaamste extensies is zonder meer op zijn plaats.

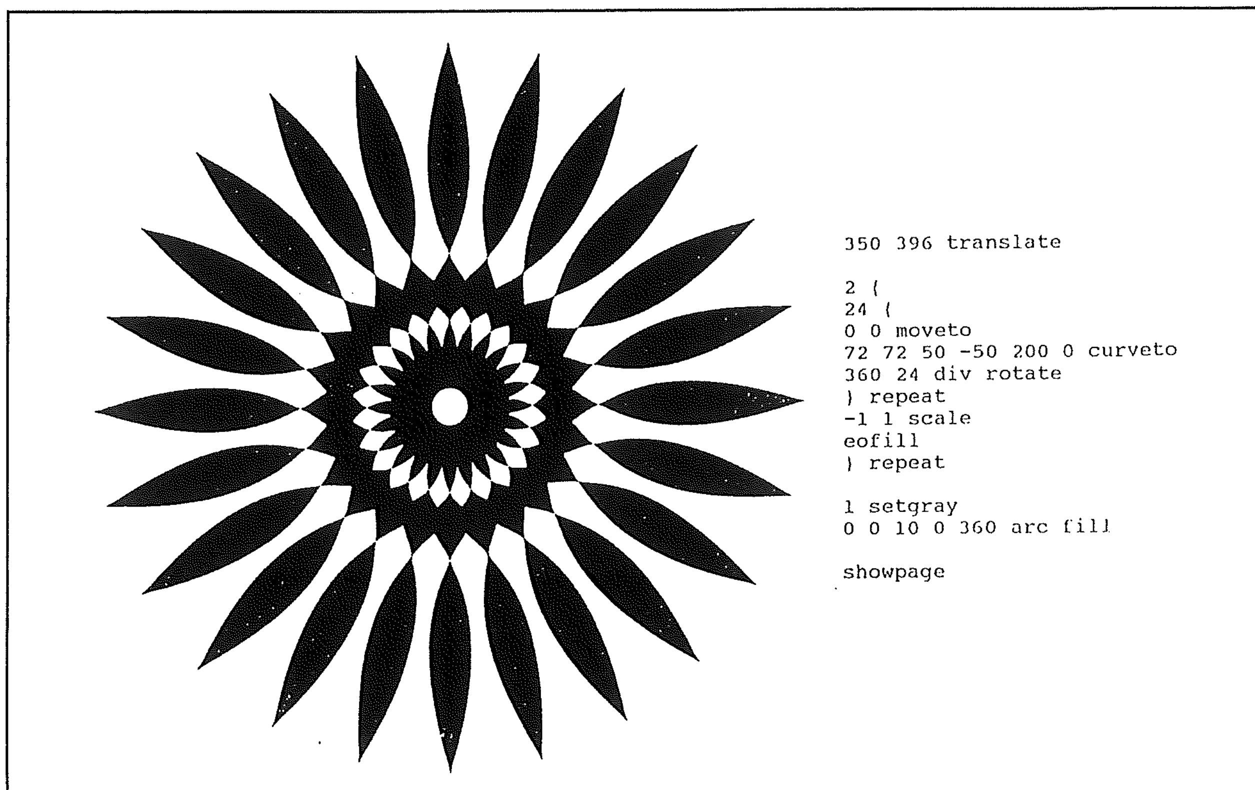
- Forms
  Vaak terugkerende gegevens in teksten, grafieken of afbeeldingen (bijvoorbeeld een logo dat op iedere pagina wordt afgedrukt) worden eenmalig door de RIP omgerekend naar een printerbeeld en kunnen als dusdanig in een cachegeheugen worden opgeslagen. Het zal duidelijk zijn dat de verwerkingssnelheid hierdoor flink wordt opgevoerd.
- Patterns
  Ingewikkelde vulpatronen worden op dezelfde manier behandeld, eenmalig omgerekend en opgeborgen in cache.



Figuur 3/19.11-61: Alle features van Postscript Level-2 overzichtelijk samengevat.

- Device-independent Color
  Met deze extensie beschikt Postscript Level-2 over de mogelijkheid rekening te houden met de specifieke eigenschappen van een kleurenprinter. Deze eigenschappen worden bij het berekenen van de CYMK-rasters in de wiskunde betrokken, zodat het in principe zo is, dat een kleurenafdruk op een goedkope 300 dpi inkjet printer er (wat kleurtinten betreft) net zo uitziet als de definitieve afdruk op een heel dure dye sublimation printer.
- Compression/Decompression
  Level-2 heeft geen moeite met het verwerken van gecomprimeerde bestanden volgens de standaarden RLE, CCITT, LZW, JPEG en ASCII-85.

19.11 Principes en werking van printers



Figuur 3/19.11-62: Het prachtige resultaat van een Postscript programmaatje, bestaande uit slechts 13 regels code.

Omvangrijke illustraties, die bijvoorbeeld met JPEG gecodeerd zijn om de omvang tot hanteerbare grootte te reduceren, kunnen rechtstreeks aan een Level-2 printer aangeboden worden.

- New Halftoning
  Level-2 bevat zeer uitgebreide algoritmen voor het berekenen van grijs- en kleurenrasters. De rasterfrequentie kan tot op 0,01 lpi gedefinieerd worden, de rasterhoek tot op 0,0001°. Op kleurgebied worden FM-rasters ondersteund.
- Performance Improvements
  Level-2 is uitgerust met een zeer efficiënt werkend systeem van geheugenbeheer, waardoor de vele MB's geheugen die in een Postscript Level-2 printer noodzakelijk zijn, zo efficiënt mogelijk worden gevuld.
- Improved Device Specific Features
  Level-2 is in staat bepaalde apparaatafhankelijk specificaties volledig te ondersteunen.
- ATM Font Rendering
  ATM, de Type-1 implementatie voor Windows, wordt door Level-2 volledig ondersteund en kan met Postscript instructies op alle mogelijke manieren bewerkt worden.

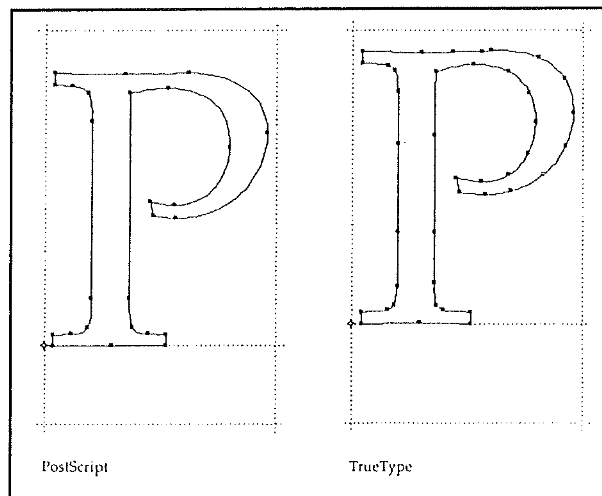
Postscript werkt, net zoals PreScribe, met BASIC-achtige commando's die als pure ASCII-file ingevoerd moeten worden. De commando-structuur van Postscript is echter veel ingewikkelder dan deze van PreScribe. Over het rechtstreeks pro-

grammeren van een Postscript printer zijn hele dikke boeken vol geschreven!
De syntax van PS-instructies is bovendien veel ingewikkelder. Wat te denken van sleutelwoorden als:

- SETSOFTWAREIOMODE;
- DOSTARTPAGE;
- SETDEFAULTTIMEOUTS.

Wie echter deze taal goed beheerst kan prachtige plaatjes op het papier krijgen. Plaatjes, die gemaakt met een grafisch programma uren ontwerpwerk kosten, kunnen met enige Postscript-regeltjes op het papier verschijnen. Als voorbeeld van de zeer krachtige Postscript instructies is in figuur 3/19.11-62 een heel klein PS-programmaatje gegeven, met het uitgeprinte resultaat.
Dergelijke resultaten kunnen natuurlijk alleen met een dynamische taal als Postscript bereikt worden, waar men bijvoorbeeld REPEAT-lussen in kan verwerken.

TrueImage
TrueImage is een PDL printertaal die in 1990 werd ontwikkeld door MicroSoft en Apple. De voornaamste bedoeling van deze gezamenlijke inspanning van twee aartsconcurrenten was de absolute hegemonie van Adobe's Postscript te doorbreken. Nu, acht jaar later, is wel duidelijk dat dit niet gelukt is. Terwijl Postscript overal springlevend is, is de toepassing van TrueImage beperkt gebleven tot MicroSoft's eigen Windows 3.1 en 95.
TrueImage is compatibel met Level-1 van Postscript en bevat bovendien veel features die eerst in Level-2 van Postscript geïmplementeerd werden. Zo ondersteunt TrueImage de Type-1 en Type-3 fonts van Postscript, maar biedt bovendien een eigen derde fontformaat aan: het van Windows zeer bekende TrueType. Ook TrueType beschouwt fonts als objecten, die dus volledig door software gemanipuleerd kunnen worden. Er zijn echter toch principiële verschillen in de manier waarop de Type-1 en de TrueType letters worden opgebouwd. Aan de hand van figuur 3/19.11-63 kunnen deze verschillen verklaard worden.
Postscript maakt gebruik van Bézier-curves om de outline van een letter te definiëren. Bij TrueType worden B-Spline tweede orde curves gebruikt. Deze curves zijn veel minder veelzijdig, met als gevolg dat de outline van een TrueType-letter in veel meer curves moet worden opgesplitst. De omvang van een TrueType font-bestand is dus groter. Maar als voordeel kan vermeld worden dat de software veel minder moet rekenen om de B-Spline curves op papier te zetten, waardoor het printen onder TrueImage veel sneller gaat dan onder Postscript.



Figuur 3/19.11-63: De letteropbouw bij Type-1 (links) met Bézier en van TrueType (rechts) met B-Spline.

Een ander voordeel van TrueImage is dat het systeem vanaf het begin ontworpen is met WYSIWYG in het achterhoofd. De

software, die een TrueImage file naar de printer stuurt, wordt ook gebruikt om het beeld op de monitor op te bouwen.
TrueImage levert dus absolute compatibiliteit tussen een ontwerp op het scherm en de afdruk op papier.
Een groot nadeel van Postscript is dat de extra hardware die in een Postscript-printer is ingebouwd het apparaat toch nog steeds gemiddeld *f* 1.000,00 duurder maakt een exacte soortgenoot zonder deze voorzieningen. Met deze goedkope printer kan men echter wel zonder meer TrueImage verwerken!

GDI
GDI is in feite geen printertaal, maar een interface die standaard bij Windows 3.1 wordt geleverd. Deze "Graphic Device Interface" GDI.EXE wordt bij het opstarten van Windows in het PC-geheugen geladen.
Applicaties kunnen vanaf dat moment de GDI-functies aanroepen om bijvoorbeeld een grafisch object op het scherm te tekenen of af te drukken op de printer. De GDI zorgt ervoor dat deze functies worden vertaald naar commandoreeksen, die begrijpelijk zijn voor de elektronica in de grafische kaart en in de printer. De GDI weet welke grafische kaart en welke printer aanwezig zijn, dank zij de device-drivers (****.DRV) die men bij de installatie van Windows heeft opgegeven.
Nu worden er bij de vertaalslag naar de printer in feite veel te veel handelingen verricht.
Het zou allemaal veel sneller kunnen als er printers waren, die de functies van de GDI rechtstreeks zouden kunnen interpreteren. Dergelijke printers zijn op de markt gekomen en zijn te herkennen aan het Windows-logo dat zij van MicroSoft mogen voeren.
Men kan echter nog een stap verder gaan. Moderne PC-systemen beschikken over snelle processoren en minstens 4 MB geheugen. Waarom zou men dan nog een extra processor en extra geheugen in de printer nodig hebben? Er zijn inmiddels zeer goedkope laserprinters op de markt verschenen, die het volledig zonder eigen intelligente elektronica of geheugen moeten stellen. De printer "leent" die hardware van het PC-systeem als er geprint wordt. Ook nu zorgt de GDI van Windows er voor, samen met een eenvoudig stuurprogramma, dat het systeem een deel van zijn werkgeheugen als printergeheugen beschouwt en de af te drukken gegevens op inktdot-niveau in dit geheugen worden opgebouwd. Nadat een volledige pagina verwerkt is, worden deze gegevens naar het kale printmechanisme van de printer gestuurd.
MicroSoft heeft inmiddels het "Windows Printing System" op de markt gebracht, een software-pakket dat GDI-printers op een heel gebruikersvriendelijke manier aanstuurt.
De rechtstreekse koppeling van printers aan de GDI van Windows heeft uiteraard een aantal voordelen.

- Snelheid
  Omdat een heleboel rekenintensieve bewerkingen vervallen neemt de printsnelheid onder Windows behoorlijk toe.
- Prijs
  Een heleboel elektronica die in normale printers aanwezig is, kan achterwege blijven bij GDI-printers. Deze maken immers gebruik van de hardware van het PC-systeem.
- Comfortabel
  Omdat een GDI-printer volledig door het Windows Printing System wordt bestuurd, worden foutmeldingen van de

19.11 Principes en werking van printers

printer in begrijpelijke taal op het beeldscherm in het WPS-venster geschreven.

GDI-printers hebben echter ook een aantal nadelen, die als zeer beperkend kunnen worden beschouwd en de marktpenetratie zeer zeker in de weg staan.

- MS-DOS besturing onmogelijk
  Voornaamste nadeel is dat deze printers aan Windows vastzitten en zonder deze schil niets op papier kunnen zetten! Rechtstreeks printen vanuit MD-DOS is dus niet mogelijk!
- Hoe waardevast blijft GDI?
  Niemand kan garanderen dat de GDI, die nu in gebruik is door Windows 3.1 en Windows 95, een constante factor is bij alle toekomstige nieuwe versies van Windows.
  Als MicroSoft zijn GDI laat vallen, dan kan men in principe niets meer beginnen met GDI-printers.

3/19.12

Principes van DMA, Direct Memory Access

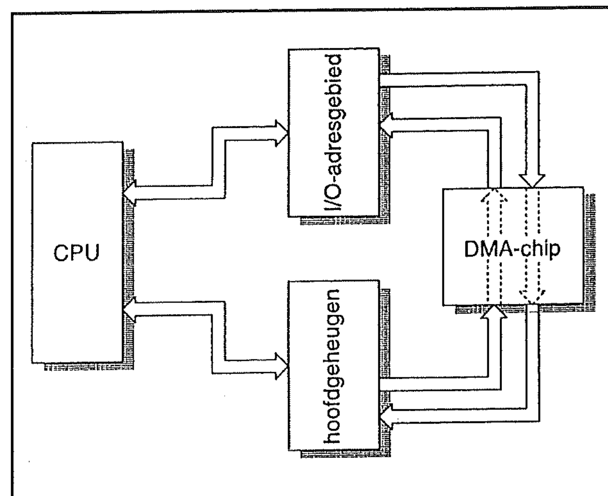
Inleiding

Wat is DMA?

DMA is het letterwoord van "Direct Memory Access", oftewel "directe geheugen benadering". Het is een systeem dat is ontwikkeld om de programmeur rechtstreeks toegang te geven tot het geheugen van een computer. Behalve de processor is er in iedere computer dus nog een IC dat zelfstandig het werkgeheugen en de periferie-eenheden kan benaderen, namelijk de DMA-chip of DMA-controller.
Wat betreft de historie kan men stellen dat bij de op Intel-processoren gebaseerde computers, dus de PC, eerst bij de ontwikkeling van de EISA-standaard het principe van DMA interessant is geworden. Daar werd namelijk voor het eerst een 32 bit chip als DMA-controller ingezet.

Directe geheugenbenadering met periferie en geheugen

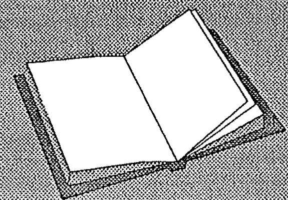
Vooral bij krachtige computers met multitasking-besturingssystemen wordt het DMA-concept gebruikt, en dit voornamelijk om de processor niet teveel te belasten met gegevensuitwisseling. De DMA-controller vormt een tweede verbinding tussen periferie en werkgeheugen in een computersysteem. In figuur 3/19.12-1 is het principe van DMA schematisch voorgesteld.



Figuur 3/19.12-1: Het principe achter DMA.

Hoe gaat geheugen-overdracht via de processor? De processor haalt de gegevens uit een perifere eenheid en zet deze in een intern register en transporteert deze vervolgens vanuit dit register naar het werkgeheugen. Dit werkt uiteraard ook in omgekeerde richting. De pijlen geven al aan,

LEES OOK:
Hoofdstuk 3/19.1



19.12 Principes van DMA, Direct Memory Access

dat het altijd een tweetraps proces is, waarbij altijd een intern processorregister is betrokken. Dit simpele werk kan bij grotere gegevensblokken de processor nogal lang bezighouden.
Hoe gaat het via DMA? De oplossing ligt besloten in de DMA-chip, die naast de processor een tweede verbinding vormt tussen periferie en werkgeheugen. De gegevens worden daarbij niet uitgewisseld via een intern register van de DMA-chip, maar direct via de databus tussen periferie en werkgeheugen. De DMA-controller levert alleen de signalen voor adressering en bussturing. De periferie-eenheid kan dus direct naar het werkgeheugen schrijven en van daaruit lezen. De processor hoeft zich niet meer bezig te houden met deze gegevensuitwisseling en kan zich bezig houden met andere klussen.

Samenvatting
Interruptbesturing is een vorm van "achtergrondacquisitie" omdat het hoofdprogramma geen code bevat om gegevens vanuit het randapparaat in te lezen. Integendeel, de processor wordt tijdens het uitvoeren van het hoofdprogramma van tijd tot tijd onmerkbaar "gestolen" om deze functie uit te voeren.
Bij DMA leest een hiervoor bestemde gegevensoverdrachtcomponent de binnenkomende gegevens van een apparaat in en slaat deze gegevens op in een geheugenbuffer van het systeem zodat de processor deze gegevens op een later tijdstip kan binnenhalen. Dit DMA-proces vindt, gezien vanuit de processor, transparant (=ongemerkt) plaats.

Voordelen van DMA
DMA biedt diverse voordelen ten opzichte van polling en interruptbesturing. DMA is snel omdat de DMA-controller de gegevens van de ene computerlocatie naar een andere verplaatst, waarvoor per keer slechts één of twee bus lees/schrijfcycli nodig zijn.
Daarnaast is DMA meestal nodig om de maximale gegevensoverdrachtsnelheid te kunnen bereiken en daarom heel nuttig voor apparatuur met zeer snelle data acquisitie. DMA beperkt tevens de wachttijd die geldt bij de afhandeling van een data acquisitie apparaat omdat de hiervoor bestemde hardware sneller reageert dan interrupts.
Ook de overdrachtstijd is kort. Door het terugbrengen van de wachttijd kan de omvang worden verminderd van het tijdelijk geheugen waarover een I/O-apparaat moet beschikken. Daarnaast ontlast DMA de processor, hetgeen betekent dat de processor geen enkele instructie hoeft uit te voeren om gegevens te verzenden. Dit houdt in dat de processor niet wordt gebruikt voor het afhandelen van de gegevensoverdracht en dus beschikbaar is voor andere taken. In systemen waarbij de processor hoofdzakelijk buiten de cache om werkt, vindt de gegevensoverdracht in feite parallel plaats hetgeen de optimale benutting van het systeem ten goede komt.

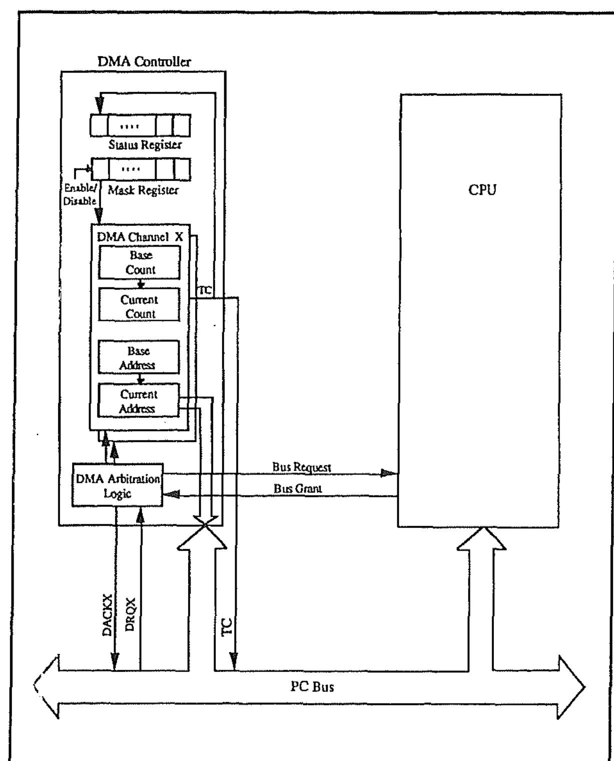
DMA-kanalen
Aan de perifere eenheden wordt meestal een DMA-kanaal toegewezen, dat door een eenheid wordt geactiveerd via een DMA-request (DREQ).
De DMA-controller reageert op dit verzoek en voert de gegevensuitwisseling uit via dit kanaal, bijvoorbeeld bij het inlezen van een diskette-sector. De processor initialiseert de DMA-sturing op de gewenste manier en geeft de bijbehorende opdracht. De diskette-controller plaatst de kop boven de gewenste sector, activeert de leeskop en de DMA-sturing. De DMA-

19.12 Principes van DMA, Direct Memory Access

controller plaatst dan de sectorgegevens in een buffer in het werkgeheugen. Door een hardware-interrupt kan de floppy-controller FDC aan de processor melden dat de leesopdracht is uitgevoerd. De complete handeling tussen opdracht aan de floppy-controller en de hardware-interrupt van uit de FDC wordt geregeld door FDC- en DMA-sturing, zonder tussenkomst van de processor.

De DMA-controller

Een DMA-controller is een IC dat zich meestal buiten de processor van de computer bevindt en is geprogrammeerd om een reeks gegevens-overdrachten uit te voeren namens de processor. Een DMA-controller heeft rechtstreeks toegang tot het geheugen en wordt gebruikt om gegevens van de ene geheugenlocatie over te brengen naar een andere, of van een I/O-apparaat naar het geheugen en vice versa.
Een DMA-controller bestuurt een aantal DMA-kanalen. Elk kanaal kan worden geprogrammeerd om één reeks DMA-overdrachten uit te voeren.
Apparaten, meestal I/O-apparaten, die gegevens verzamelen welke moeten worden gelezen (of apparaten die gegevens moeten uitvoeren en waarnaar moet worden geschreven) laten de DMA-controller via het sturen van een hardware DMA-verzoeksignaal (DMA-request) weten dat er een DMA-overdracht moet worden uitgevoerd. Een DMA-verzoeksignaal voor elk kanaal wordt naar de DMA-controller geleid. Dit signaal wordt kritisch gevolgd en er wordt op gereageerd op bijna dezelfde manier waarop een processor de interrupts verwerkt. Wanneer de DMA-controller een DMA-verzoek ontdekt, reageert hij hierop door één of meer gegevensoverdrachten uit te voeren van dat I/O-apparaat naar het systeemgeheugen of omgekeerd. De kanalen moeten door de processor worden vrijgegeven zodat de DMA-controller kan reageren op DMA-verzoeken. Het aantal overdrachten dat kan worden uitgevoerd, de gebruikte overdrachtsmodi en de aangesproken geheugenlocaties zijn afhankelijk van de manier waarop het DMA-kanaal is geprogrammeerd.
Een DMA-controller deelt het systeemgeheugen en de I/O-bus met de processor en kan zowel als bus master (meester) of als slave (slaaf) fungeren. In figuur 3/19.12-2 is de architectuur van de DMA-controller afgebeeld en de manier waarop deze samenwerkt met de processor. In de busmeester bedrijfstoestand krijgt de DMA-controller van de processor de systeembus (adres-, data- en besturingslijnen) toegewezen om DMA-overdrachten uit te voeren.



Figuur 3/19.12-2: De interne architectuur van een DMA-controller.

Omdat de processor de systeembus vrijgeeft voor de duur van de overdracht, werd deze procedure ook wel aangegeduid als "cycle stealing". Deze term is echter met het oog op de nieuwe, zeer krachtige architecturen die in PC's voorkomen, niet meer zo van toepassing. Deze architecturen maken gebruik van cachegeheugen voor de processor, waardoor de DMA-controller tot op zekere hoogte gelijktijdig met de processor in bedrijf kan zijn.
In de busslaaf bedrijfstoestand krijgt de processor toegang tot de DMA-controller. De processor programmeert nu de interne registers van de DMA-controller zodat er DMA-overdrachten kunnen plaatsvinden. De interne registers omvatten de bron- (source) en bestemmings- (destination) adresregisters en de overdracht telregisters (transfer count registers) voor elk DMA-kanaal, naast besturings- (control) en statusregisters voor het starten, kritisch volgen en gaande houden van de werkzaamheden van de DMA-controller.

DMA-signalen

Inleiding
Bij een DMA-cyclus worden nogal wat signalen geactiveerd en het is van groot belang om te weten wélke signalen in het systeem bij een dergelijke actie betrokken zijn en wat zij precies doen.

$\overline{\text{IOR}}$ (I/O Read)
Als de DMA-controller is uitgeschakeld, kan de processor een intern register van de DMA-chip lezen door $\overline{\text{IOR}}$ met een laag signaal te activeren. Is de DMA-chip actief en stuurt hij de data- en adresbus, dan geeft een actief $\overline{\text{IOR}}$-signaal aan dat de DMA-chip via een poort-adres data leest van een perifere eenheid. Het signaal $\overline{\text{IOR}}$ kan ook actief zijn bij het schrijven van periferie naar geheugen.

$\overline{\text{IOW}}$ (I/O Write)
Als de DMA-chip is uitgeschakeld, kan de processor een intern register van de DMA-chip schrijven door $\overline{\text{IOR}}$ met een laag signaal te activeren. Is de DMA-chip actief en stuurt hij de data- en adresbus, dan geeft een actief $\overline{\text{IOW}}$-signaal aan dat de DMA-chip via een poort-adres data schrijft naar een perifere eenheid. Het signaal kan ook actief zijn bij het lezen van geheugen naar periferie.

$\overline{\text{MEMR}}$ (Memory Read)
Met behulp van het $\overline{\text{MEMR}}$-signaal meldt de DMA-controller aan de bussturing, dat er gegevens uit het werkgeheugen worden gelezen. Het $\overline{\text{MEMR}}$-signaal kan ook actief (laag) zijn bij het lezen, of bij het kopiëren van gegevens binnen het werkgeheugen.

$\overline{\text{MEMW}}$ (Memory Write)
Met het $\overline{\text{MEMW}}$-signaal meldt de DMA-controller aan de bussturing, dat er gegevens naar het werkgeheugen moeten worden geschreven. Het signaal kan ook actief (laag) zijn bij het schrijven, of bij het kopiëren van gegevens binnen het werkgeheugen.

READY
Langzame geheugens en periferie-eenheden kunnen het ingangssignaal READY activeren om de schrijf- en leescycli van de DMA-controller te verlengen.

HLDA (Hold Acknowledge)
Een HLDA-signaal vanuit de processor of een andere busmaster geeft aan, dat de

processor de lokale bus heeft vrijgegeven en dat de DMA-controller de controle kan krijgen voor de gegevens-overdracht.

ADSTB (Address Strobe)
Dit signaal wordt gebruikt om de hoge adresbytes A8 tot en met A15 in een externe DMA-adreslatch in te lezen. De DMA-controller activeert het ADSTB-signaal, zodra deze bytes klaarstaan op de databus DB0-DB7.

AEN (Address Enable)
De DMA-controller activeert via dit signaal de externe DMA-adreslatch.
Als AEN hoog is, zet de DMA-adreslatch het opgeslagen adres als adresbits A8-A15 op de adresbus. AEN kan ook worden gebruikt om andere busdrivers uit te schakelen die ook de bits A8-A15 genereren of om periferie-eenheden uit te schakelen.

HRQ (Hold Request)
De DMA-controller vraagt via HRQ of hij de sturing van de lokale bus kan overnemen van de processor of een andere busmaster.
De processor reageert met een HLDA-signaal. De DMA-controller activeert HRQ zodra een niet-gemaskeerde DRQx of een software-gestuurde DMA-request optreedt.

$\overline{\text{CS}}$ (Chip Select)
De processor activeert $\overline{\text{CS}}$ om de interne registers van de DMA-controller te kunnen benaderen voor het lezen of schrijven van opdracht- en databytes. De gegevensuitwisseling met de processor geschiedt via de databus DB0-DB7. Bij DMA-transfers wordt de $\overline{\text{CD}}$-ingang automatisch intern gedeactiveerd door de DMA-controller.

CLK (Clock)
Deze ingang ontvangt het DMA-kloksignaal en bestuurt het gehele proces. De DMA-clock wordt afgeleid van de hoofdclock van de processor.

RESET
Een hoog signaal op deze pen initialiseert de DMA-controller opnieuw.

DACK0 tot en met DACK3 (DMA Acknowledge)
Als een DACKx-signaal actief is, is het bijbehorende DMA-kanaal actief en wordt de perifere eenheid afgehandeld die via DRQx een DMA-request heeft gedaan. Op elk willekeurig tijdstip kan er slechts één DACKx actief zijn.
De DMA-controller activeert een DACKx-leiding pas nadat hij met behulp van HRQ en HLDA de lokale bussturing heeft overgenomen. De polariteit van de signalen (laag- of hoog-actief) kan worden vastgelegd in het modusregister.

DREQ0 tot en met DREQ3 (DMA Request)
Een actief signaal van een perifere eenheid aan een pen DREQx geeft aan, dat de eenheid een DMA-transfer aanvraagt. Een diskette-controller kan bijvoorbeeld een DREQx-leiding activeren om de gelezen gegevens te kunnen overdragen naar het werkgeheugen. Meestal heeft DREQ0 de hoogste en DREQ3 de laagste prioriteit. De request wordt bevestigd met het overeenkomstige DACKx-signaal van de DMA-controller. Ook de polariteit van deze signalen (laag- of hoog-actief) kan worden vastgelegd in het modusregister.

DB0 tot en met DB7 (Data Bidirectional)
Deze acht pennen vormen de bidirectionele databus van de DMA-controller bij

lees- en schrijfhandelingen voor interne registers of bij DMA-transfers. In DMA-cycli worden via DB0 tot en met DB7 de hoge acht bits van het DMA-adres doorgegeven die via ADSTB in de externe DMA-adreslatch worden geladen. Bij interne geheugentransfers wordt het door te geven databyte via DB0-DB7 eerst in het interne register geladen en dan vanuit het register via DB0-DB7 weer uitgevoerd.

A0 tot en met A3
Deze signalen vormen een bidirectioneel adres-nibble. Als de DMA-controller niet actief is kan de processor via dit nibble interne registers adresseren.
Is de DMA-controller actief, dan worden de vier laagste adresbits uitgevoerd via A0-A3.

A4 tot en met A7
Deze vier signalen vormen het tweede adres-nibble als de DMA-controller actief is. Is de DMA-controller niet actief, dan zijn deze pennen uitgeschakeld.

$\overline{\text{EOP}}$ (End of Process)
Dit bidirectionele signaal geeft de voltooiing aan van een DMA-transfer. Als de tellerstand van het actieve kanaal in de DMA-controller de waarde 0 heeft bereikt, geeft de DMA-controller een actief (laag) $\overline{\text{EOP}}$-signaal om de aan perifere eenheid het einde van de overdracht te melden. Anderzijds kan de perifere eenheid $\overline{\text{EOP}}$ laten afvallen, om de DMA-chip te melden dat de overdracht voortijdig is beëindigd. Dat is bijvoorbeeld het geval, wanneer een buffer in de perifere eenheid door de DMA-controller is geleegd en de gegevens volledig naar het werkgeheugen zijn overgedragen. Zowel door de interne als de externe $\overline{\text{EOP}}$-conditie wordt het TC-bit in het statusregister op 1 gezet, het bijbehorende request-bit op 0 gezet en de DMA-transfer beëindigd.

Het werken met DMA

Inleiding
In dit hoofdstuk wordt uiteengezet hoe DMA wordt uitgevoerd in een gangbare PC-architectuur, waarbij diverse PC/DMA-implementaties met elkaar worden vergeleken. Ook de software die nodig is voor het gebruik van DMA bij data acquisitie met de computer komt hier aan de orde.

Maar ...
Maar allereerst is enige kennis over de diverse DMA-overdrachttypen en -modi absoluut noodzakelijk! Toegegeven, zware kost, maar onontbeerlijk voor het begrijpen van het verschijnsel "DMA".

DMA-overdracht typen

Inleiding
DMA-controllers verschillen van elkaar al naar gelang het soort DMA-overdracht en het aantal DMA-kanalen dat ze ondersteunen.
Er zijn twee typen DMA-overdracht, te weten:
- flyby;
- fetch-and-deposit.

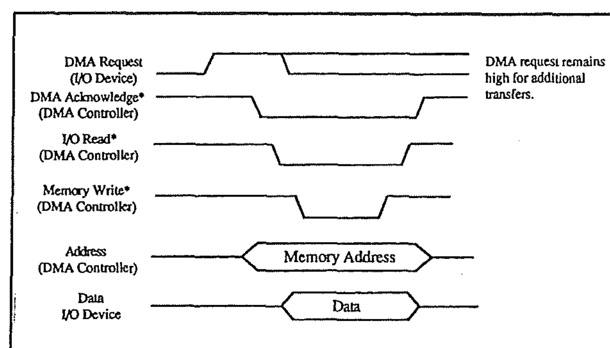
De drie gangbare overdrachtsmodi zijn:
- single;
- block;
- demand.

De beide typen DMA-overdracht en de verschillende modi worden hieronder toegelicht.

19.12 Principes van DMA, Direct Memory Access

Flyby DMA-overdracht
Het snelste type DMA-overdracht wordt aangeduid als enkelvoudige cyclus (single-cycle), enkelvoudig adres (single-address) of "flyby" overdracht. Bij een flyby DMA-overdracht wordt één busbewerking gebruikt om de overdracht uit te voeren zodat de gegevens van de bron worden gelezen en gelijktijdig naar de bestemming worden geschreven. Voor een flyby overdracht zet het apparaat dat de dienstverlening aanvraagt een DMA-verzoek op de daarvoor bestemde kanaalverzoeklijn (channel request line) van de DMAcontroller. De DMA-controller reageert hierop door de besturing van de systeembus over te nemen van de processor en daarna het voorgeprogrammeerde geheugenadres af te geven. Tegelijkertijd stuurt de DMA-controller een DMA-bevestigingssignaal (DMA acknowledge) naar het aanvragende apparaat. Dit signaal maakt het aanvragende apparaat erop attent dat de gegevens op de databus van het systeem kunnen worden gezet of dat de gegevens van de systeembus kunnen worden afgehaald, al naar gelang de overdrachtsrichting. Met andere woorden: een flyby DMA-overdracht ziet er uit als een geheugen lees- of schrijfcyclus waarbij de DMA-controller het adres levert en het I/O-apparaat de gegevens leest of schrijft. Omdat met een flyby DMA-overdracht slechts één geheugencyclus per gegevensoverdracht is gemoeid, is dit type DMA-overdracht uiterst efficiënt, hoewel geheugen-naar-geheugen-overdracht daarbij niet mogelijk is. De flyby DMA-overdracht is in figuur 3/19.12-3 schematisch weergegeven.

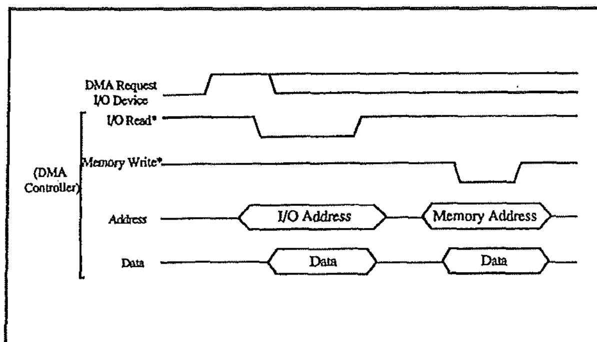
Fetch-and-deposit DMA-overdracht
Het tweede type DMA-overdracht wordt aangeduid als dubbele cyclus (dual-cycle), dubbel adres (dual-address), flow-through of "fetch-and-deposit" DMA-overdracht. Zoals de benaming al aangeeft, zijn bij dit type overdracht twee geheugen- of I/O-cycli betrokken. De gegevens die moeten worden verzonden, worden eerst uit het I/O-apparaat of het geheugen gelezen en tijdelijk opgeslagen in een intern dataregister van de DMA-controller. De gegevens worden nadien in de volgende cyclus naar het geheugen of naar een I/O-apparaat geschreven. In figuur 3/19.12-4 is de 'fetch-and-deposit' DMA-overdracht schematisch weergegeven. Hoewel deze methode niet zo efficiënt is omdat de DMA-controller twee cycli uitvoert en daardoor de systeembus langer bezet houdt, is dit type overdracht geschikt voor het koppelen van apparaten met verschillende databusbreedten. Zo kan een DMA-controller bijvoorbeeld twee 16 bit brede leesopdrachten van de ene locatie uitvoeren, gevolgd door een 32 bit brede schrijfopdracht naar een andere locatie.



Figuur 3/19.12-3: Timing bij flyby DMA-overdracht.

Een DMA-controller die dit type overdracht ondersteunt, beschikt naast de gebruikelijke tel- en besturingsregisters voor de overdracht over twee adresregisters per

kanaal (bronadres en bestemmingsadres) en registers voor de busbreedte. In tegenstelling tot een flyby-bewerking is dit type DMA-overdracht geschikt voor zowel geheugen-naar-geheugen als I/O-overdrachten.



Figuur 3/19.12-4: Fetch-and-deposit DMA-overdracht.

Overdrachtsmodi

Inleiding

Naast de beide typen DMA-overdracht hebben DMA-besturingseenheden één of meer DMA-overdrachtsmodi:
- single;
- block;
- demand;

zijn de meest gangbare overdrachtsmodi.

Single

De enkelvoudige (single) overdrachtsmodus verzendt één datawaarde voor elke melding van een DMA-verzoek. Deze modus is de langzaamste overdrachtsmethode omdat voor elke overdracht de DMA-controller als "scheidsrechter" voor de systeem-bus moet fungeren.
Deze arbitrage is geen al te groot probleem wanneer de bus slechts licht is belast maar kan leiden tot wachttijdproblemen wanneer meerdere apparaten gebruik maken van de bus.

Block en demand

De blok (block) en aanvragende (demand) overdrachtsmodi verhogen de doorvoersnelheid van het systeem omdat ze de DMA-controller in staat stellen om meerdere DMA-overdrachten uit te voeren wanneer deze de bus toegewezen heeft gekregen. In de blokoverdrachtsmodus voert de DMA-controller als antwoord op een enkelvoudig DMA-verzoek van het I/O-apparaat met de hoogst mogelijke snelheid de volledige DMA-reeks uit zoals die is gespecificeerd door het overdracht-telregister (transfer count register). In de "demand mode" overdracht voert de DMA-controller de DMA-overdrachten met de hoogst mogelijke snelheid uit zo lang het I/O-apparaat zijn DMA-verzoek handhaaft. Wanneer het I/O-apparaat zijn DMA-verzoek intrekt, wordt de overdracht stopgezet.

Werking van de DMA-controller

Inleiding

Voor elk kanaal slaat de DMA-controller het geprogrammeerde adres en het aantal te verzenden bytes (count) op in de basis- (base) registers. Kopieën van de informatie worden in de huidige adres- (current address) en de huidige telregisters (current count) bewaard.
Elk DMA-kanaal wordt vrijgegeven en geblokkeerd via een DMA-maskeer (mask) register.
Wanneer DMA wordt gestart via een schrijfopdracht naar de basisregisters en

19.12 Principes van DMA, Direct Memory Access

door het DMA-kanaal vrij te geven, worden de huidige registers geladen vanuit de basisregisters. Bij elke DMA-overdracht wordt de waarde in het huidige adresregister op de adresbus gezet en wordt het huidige adresregister automatisch opgehoogd of afgelaagd. Het huidige telregister bepaalt het aantal overdrachten dat resteert en wordt automatisch afgelaagd na elke overdracht.

Terminal count

Wanneer de waarde in het huidige telregister van 0 naar -1 gaat, wordt er een "terminal count" (TC) signaal opgewekt om aan te geven dat de DMA-overdrachtreeks is voltooid. Met deze actie wordt te kennen gegeven dat de "terminal count" is bereikt.
DMA-controllers genereren vaak een hardware TC-impuls tijdens de laatste cyclus van een DMA-overdrachtreeks. Dit signaal kan kritisch worden gevolgd door de I/O-apparaten die deelnemen aan het DMA-verkeer.
DMA-controllers moeten opnieuw worden geprogrammeerd wanneer een DMA-kanaal een TC bereikt. Dientengevolge nemen DMA-controllers wat processortijd in beslag, maar aanzienlijk minder dan de processor nodig zou hebben om I/O-interrupts van de apparatuur te verwerken. Wanneer een DMA-kanaal een TC bereikt, kan het nodig zijn dat de processor de besturingseenheid opnieuw programmeert voor aanvullende DMA-overdrachten. Sommige DMA-controllers onderbreken de processor telkens wanneer een kanaal afschakelt. DMA-controllers beschikken ook over bepaalde mechanismen waardoor, na voltooiing van de DMA-overdrachtreeks, een DMA-kanaal automatisch opnieuw wordt geprogrammeerd. Tot deze mechanismen behoren onder andere de auto-initialisatie en het koppelen van buffers (buffer chaining).
Ten gevolge van de auto-initialisatie wordt de DMA-overdrachtreeks herhaald door aan het eind van een DMA-reeks de huidige registers van het DMA-kanaal opnieuw te laden vanuit de basisregisters en het kanaal opnieuw vrij te geven. Bufferkoppeling wordt toegepast bij de overdracht van gegevensblokken naar niet-aangrenzende buffergebieden of bij het verwerken van dubbel-gebufferde data acquisitie. Bij bufferkoppeling onderbreekt een kanaal de processor en wordt dit kanaal geprogrammeerd met het volgende adres en de telparameters, terwijl DMA-overdrachten worden uitgevoerd via de huidige buffer. Sommige DMA-controllers beperken de tussenkomst van de processor nog meer doordat ze over een ketting-adresregister beschikken dat wijst naar een ketting-controletabel in het geheugen. De DMA-controller laadt dan zijn eigen kanaalparameters vanuit het geheugen. Over het algemeen geldt dat naarmate de DMA-controller meer geperfectioneerd is, de processor minder "tussenbeide" hoeft te komen.

Statusregisters

Een DMA-controller heeft één of meer statusregisters die door de processor worden gelezen om de toestand van elk DMA-kanaal te bepalen. Het statusregister geeft aan of er een DMA-verzoek is uitgegaan op een bepaald kanaal en of een kanaal de TC heeft bereikt.
Door het lezen van het statusregister wordt vaak de "terminal count"-informatie in het register gewist, hetgeen tot problemen leidt wanneer meerdere programma's proberen om verschillende DMA-kanalen te gebruiken.

DMA in de PC

DMA-implementatie

Alle DMA-controllers die in dit hoofdstuk worden besproken, zijn verwant aan de Intel 8237A DMA-controller die in de originele IBM-PC wordt toegepast. Deze DMA-controller heeft vier kanalen met een 16 bit geheugenadres en bytetelregisters voor elk kanaal.
Het 16 bit telregister beperkt de lengte van een voorgeprogrammeerde DMA-overdrachtreeks tot 64 kB. De PC voegt extern een 8 bit paginaregister toe aan de DMA-controller om het 16 bit adres uit te breiden naar het 24 bit PC geheugenadres.
Ondanks deze uitbreiding kunnen voorgeprogrammeerde DMA-overdrachtreeksen de 64 kB paginagrens echter niet overschrijden omdat het paginaregister niet wordt opgehoogd wanneer het adresregister de waarde OFFFFH bereikt. De DMA-controller implementeert enkelvoudige (single), blok (block) en aanvragende (demand) DMA-overdrachtsmodi.

Twee DMA-kanalen

De PC maakt gebruik van twee van de DMA-kanalen. Het ene kanaal wordt gebruikt voor de overdracht van gegevens vanaf het diskettestation naar het geheugen en omgekeerd (kanaal 2), het andere kanaal voor geheugen-verversingscycli (kanaal 0). De overige kanalen zijn beschikbaar voor I/O-kaarten die in de uitbreidingsconnectoren van de PC kunnen worden gestoken, zoals communicatie randapparatuur. De PC I/O-kanaalbus heeft een TC signaal, DMA aanvraagsignalen DRQX en DMA bevestigingssignalen DACKX, waarbij X het DMA-kanaalnummer aangeeft.

ISA uitbreidingen voor de DMA implementatie

Bij de introductie van de 16 bit "Industry Standard Architecture" (ISA) en de PC/AT computer werd het aantal DMA-kanalen uitgebreid tot zeven. Twee Intel DMA-controllers zijn in cascade geschakeld en elke chip heeft vier kanalen. DMA-controller 1 heeft de kanalen 0 tot en met 3 en implementeert de oorspronkelijke 8 bit flyby DMA-overdrachten van de PC tussen de 8 bit I/O-adapters en het 8 bit of 16 bit systeemgeheugen.
DMA-controller 2 kreeg de kanalen 4 tot en met 7 toegewezen. Kanaal 4 wordt gebruikt voor het in cascade schakelen van de beide DMA-controllers en is daardoor niet beschikbaar voor algemeen gebruik. De kanalen 5, 6 en 7 implementeren 16 bit flyby DMA-overdrachten tussen de 16 bit I/O-adapters en het 16 bit systeemgeheugen. De telregisters in DMA-controller 2 werken als 16 bit woordtellende registers en de adressen A1 tot en met A16 worden in de adresregisters geschreven. De paginaregisters bevatten de bovenste zeven bits van het 24 bit adres en adres A0 wordt verondersteld nul te zijn. Voor een 16 bit overdracht wordt de maximale overdrachtlengte uitgebreid tot 128 kB zodat alle DMA-overdrachten binnen deze 128 kB paginagrens probleemloos kunnen worden uitgevoerd. Met een 8 MHz ISA-bus werkt de DMA-controller op 4 MHz (de klokfrequentie mag maximaal 5 MHz zijn) hetgeen neer komt op een overdrachtssnelheid van 800 kB in de aanvragende (demand) modus voor 8 bit overdrachten en een snelheid van 1,6 MB/s voor 16 bit overdrachten.

Memory refresh

Zowel bij de oorspronkelijke PC als bij de ISA computers wordt de maximale DMA-

19.12 Principes van DMA, Direct Memory Access

overdrachtssnelheid beperkt door de noodzaak om het geheugen te verversen. Een geheugenverversingscyclus (memory refresh) moet iedere 15 µs worden uitgevoerd. Een apparaat dat is ontworpen voor aanvragende (demand) DMA-overdracht moet met opzet DRQ-signalen laten schieten en elke 15 µs afstand doen van de bus zodat het geheugen kan worden ververst. In een ISA-systeem is blokoverdrachtsmodus niet zo praktisch omdat de DMA-controller pas afstand doet van de bus wanneer de gehele DMA-reeks is voltooid en er slechts 24 ISA overdrachten in 15 µs kunnen worden uitgevoerd. Aangezien langdurige DMA-overdrachten moeilijk kunnen worden geïmplementeerd en omdat meerdere apparaten gebruik maken van de DMA-controller, moeten I/O-apparaten eigenlijk voorzien zijn van eigen buffers. Dit geldt met name voor tijdkritische applicaties zoals tellerbestuurde patroonopwekking waarbij de gegevens met een vaste snelheid moeten worden uitgevoerd.

Micro Channel DMA-implementatie

Voor de Micro Channel DMA-controller is voortgeborduurd op de ISA DMA-controller, maar er zijn enkele belangrijke wijzigingen doorgevoerd. De Micro Channel heeft acht DMA-kanalen die zowel 8 bit als 16 bit gegevens kunnen verzenden. De Micro Channel maakt geen gebruik van DRQ- of DACK*-signalen. In plaats daarvan maakt de Micro Channel gebruik van de bus master arbitragesignalen voor DMA-aanvragen op basis van hetzelfde protocol dat busmasters gebruiken voor het aanvragen van de bus.
In plaats dat het I/O-apparaat de bus bestuurt, herkent de DMA-controller echter het arbitrageniveau, hij wordt busmaster en voert een "fetch-and-deposit" DMA-overdracht uit. Deze overdracht is een geheugencyclus, gevolgd door een I/O-cyclus of andersom, waarbij de gegevens tussen de cycli tijdelijk worden opgeslagen. Het I/O-apparaat kan de arbitragesignalen tijdens de I/O-cyclus gebruiken als DMA-bevestigingssignalen zodat het programmeren van elk kanaal net zo vlot verloopt als bij de ISA-programmering.
Het I/O-apparaat kan gedurende de I/O-cyclus ook worden geadresseerd door de DMA-controller.
De DMA-controller beschikt over een reeks registers die compatibel zijn met de ISA DMA-controller plus uitgebreide registers en programmeermodi. Hiervoor beschikt de DMA-controller over volledige 24 bit geheugenadresregisters (totale adresseerruimte van 16 MB) en 16 bit I/O-adresregisters. Daardoor kent de besturingseenheid geen 64 kB of 128 kB overdracht- en pagina-adresseringsbeperkingen zoals de ISA. De DMA-controller ondersteunt zowel enkele (single) als aanvragende (demand) overdrachtsmodi.

Beperkingen

De Micro Channel DMA-controller heeft enkele beperkingen die de ISA-implementatie niet kent. De auto-initialisatiemodus wordt niet ondersteund en een kanaal wordt automatisch geblokkeerd bij het bereiken van TC. Dit betekent dat de DMA-controller opnieuw moet worden geprogrammeerd zodra er een kanaal afschakelt, los van de vraag of het kanaal weer opnieuw start met dezelfde adres-en tellerwaarde. Daarnaast moet het I/O-apparaat het optreden van een TC herkennen en geen verdere DMA-aanvragen doen totdat het kanaal opnieuw is geprogrammeerd en weer is vrijgegeven. Zoniet, dan zal het systeem na een time-out periode bezwaar anteke-

nen. In de ISA-implementatie zal de DMA-controller ofwel automatisch initialiseren en doorgaan met de gegevensoverdracht, ofwel de DMA-aanvraag negeren totdat de besturingseenheid opnieuw is geprogrammeerd. Beide beperkingen maken het ontwerpen van DMA-software voor een Micro Channel I/O-apparaat moeilijker dan voor een ISA-apparaat.

Overdrachtsnelheid
De Micro Channel DMA-controller kan een maximale overdrachtssnelheid bereiken die dicht tegen de 5 MB/s aan ligt. Omdat de Micro Channel DMA-arbitrage in serie met de gegevensoverdracht plaatsvindt, zal een arbitrage-activiteit de busbandbreedte van de totale gegevensoverdracht enigszins verminderen. Net als bij de ISA-implementatie wordt elke 15 µs een geheugenverversingscyclus uitgevoerd, wat de DMA-overdrachtssnelheid ook enigszins nadelig beïnvloed. Daarnaast is het tussentijds opslaan van gegevens noodzakelijk voor tijdkritische DMA-bewerkingen.

EISA-uitbreidingen van de ISA DMA-implementatie
In plaats van een op zichzelf staand IC is de DMA-controller voor de Extended Industry Standard Architecture (EISA) bus geïntegreerd in een multifunctioneel IC. Als multifunctioneel IC maakt de besturingseenheid deel uit van de EISA chipset die zich bevindt op het moederbord van een EISA computer. Het DMA-gedeelte van het IC bestaat uit een superset van de DMA-controller die wordt gebruikt in de PC/AT (ISA) computer.

Compatibiliteit
De EISA DMA-controller is qua software compatibel met en heeft alle eigenschappen van de ISA DMA-controller, behalve dan dat de speciale geheugen-naar-geheugen overdrachtsmodus niet wordt ondersteund. Elk DMA-kanaal implementeert de enkelvoudige (single), blok (block) en aanvragende (demand) DMA-overdrachtsmodi en alle DMA-overdrachten zijn van het flyby-type. Er zijn enkele afwijkingen met betrekking tot de programmeervolgorde (met het oog op de software-compatibiliteit) en de programma's die voor EISA computers zijn geschreven maken gebruik van een andere programmeervolgorde dan de ISA-programma's.

Krachtiger
De EISA DMA-controller is aanzienlijk krachtiger dan de ISA DMA-controller. De EISA DMA-controller heeft zeven kanalen die elk de volledige 32 bit adressering (4 GB) ondersteunen, naast 32 bit gegevensoverdrachten en 24 bit byte/woord-tellers. Net als bij de Micro Channel implementatie is vanwege de volledige systeemadressering een paginaregister overbodig en hoeft ook daarna het DMA-kanaal niet opnieuw te worden geprogrammeerd om de paginagrenzen te kunnen overschrijden.
Het paginaregister is echter nog wel aanwezig met het oog op ISA-software compatibiliteit. Een 24 bit byte-telregister houdt in dat een kanaal kan worden geprogrammeerd voor het verzenden van maximaal 16 MB gegevens. Elk kanaal kan 8 bit, 16 bit of 32 bit gegevensoverdrachten uitvoeren.
De EISA DMA-controller ondersteunt eveneens een dynamische instelling van de busbreedte zodat DMA-overdrachten mogelijk zijn tussen apparaten met een verschillende databusbreedte. In het geval van een EISA DMA-controller kan een

19.12 Principes van DMA, Direct Memory Access

I/O-apparaat de TC-lijn sturen om de DMA-overdrachtreeks vroegtijdig af te breken en software-aanvragen kunnen de DMA-overdrachten besturen.
De EISA DMA-controller biedt een belangrijke toename van de overdrachtsnelheid in vergelijking tot de ISA-oplossing. Veel van deze nieuwe overdrachtsnelheden kunnen worden gebruikt door bestaande ISA I/O-kaarten.

Timing-modi

Er zijn vier verschillende timing-modi voor de DMA-overdracht:
- ISA-compatibel;
- type A;
- type B;
- Type C (burst).

De verschillende maximaal haalbare overdrachtsnelheden zijn weergegeven in figuur 3/19.12-5.

DMA-cyclustype	Overdrachtsnelheid (Mbyte/s)	Compatibiliteit
ISA-compatibel:		
8-bit	1,0	Alle ISA computers
16-bit	2,0	Alle ISA computers
Type A:	1,3	De meeste ISA computers
8-bit	2,6	De meeste ISA computers
16-bit	5,3	Alleen EISA computers
32-bit		
Type B:	2,0	Sommige ISA computers
8-bit	4,0	Sommige ISA computers
16-bit	8,0	Alleen EISA computers
32-bit		
Type C (salvo):	8,2	Alleen EISA computers
8-bit	16,5	Alleen EISA computers
16-bit	33,0	Alleen EISA computers
32-bit		

Figuur 3/19.12-5: DMA-timingmodi en -overdrachtssnelheden.

Platform	Aantal DMA-kanalen	Data formaat	Maximale buffer-grootte	Type DMA	DMA overdrachtsnelheden (Mbytes/s)		
					8-bit	16-bit	32-bit
PC/XT	4	8-bit	64 Kbyte	flyby	0,9	-	-
PC/AT (ISA)	7 (vier 8-bit, drie 16-bit)	8-bit, 16-bit	64 Kbyte of woorden	flyby	0,8	1,6	-
EISA	7	8-bit, 16-bit, 32-bit	16 Mbyte	flyby geen burst	1-2	2-4	5-8
EISA	7	8-bit, 16-bit, 32-bit	16 Mbyte	flyby burst	8,2	16,5	33,0
Micro Channel	8	8-bit, 16-bit	16 Mbyte	fetch-and-deposit	2,5	5,0	-

Figuur 3/19.12-6: Vergelijking van de capaciteiten van de diverse DMA-controllers.

Afschakelmodi

Naast de standaard scenario's "overdracht van de bufferinhoud" en "afschakelen of auto-initialiseren" biedt de EISA DMA-controller enkele handige afschakelmodi, waaronder de "buffer chaining mode" en de "ring buffer mode". In de buffer chaining modus onderbreekt de DMA-controller de processor voor het aanvragen van een nieuw bufferadres en de tellerstand. Deze onderbreking vindt telkens plaats wanneer de adresregisters worden herladen vanuit de basisregisterset. In de buffer chaining modus kunnen meerdere buffers worden gevuld/geleegd met een minimum aan processor-interventie en zonder vertraging tussen de buffers.

Platform	Geavanceerde buffer eigenschappen
PC/XT	Auto-initialisatie
PC/AT (ISA)	Auto-initialisatie
EISA	Auto-initialisatie, buffer chaining, ring buffers
Micro Channel	Niet van toepassing

Figuur 3/19.12-7: Vergelijking van de specifieke eigenschappen van DMA-controllers op het gebied van buffferbeheer.

Ring buffer modus

De ring buffer modus vormt een uitbreiding op de auto-initialisatie modus. Er kan een stopregister worden geprogrammeerd met een adres dat wordt vergeleken met het huidige adresregister. Zodra er overeenkomst wordt geconstateerd, stopt de DMA-controller met de overdrachten totdat het stopregister opnieuw is geprogrammeerd. De ring buffer modus kan bij de auto-initialisatie worden gebruikt om een circulaire buffer te beheren en om te voorkomen dat er per ongeluk gegevens worden overschreven.

Globale capaciteitsvergelijking

In de tabellen van de figuren 3/19.12-6 en -7 zijn de globale capaciteiten en specifieke eigenschappen van DMA-controllers in diverse PC-configuraties weergegeven.

DMA-software

Inleiding

Het toepassen van DMA-controllers vraagt om een zorgvuldige opzet van de software. Wat dit punt betreft zijn er twee belangrijke aspecten waarmee rekening moet worden gehouden bij het toepassen van DMA-controllers.

Het eerste aspect is dat de vele kanalen in een DMA-controller worden gebruikt door verschillende applicaties die niet met elkaar in verband staan. Omdat het een systeembron betreft die wordt gedeeld, moet de DMA-controller goed worden beheerd om te voorkomen dat de niet-gerelateerde applicaties elkaar nadelig beïnvloeden.

Ook de DMA-buffers moeten worden beheerd omdat er gegevens worden verzameld of gegenereerd. Het beheer van de buffer is sterk afhankelijk van de applicatie.

Een goed ontwerp van het DMA I/O-apparaat ondersteunt de software bij het in stand houden van betrouwbare statusinformatie en bij het besturen van de gegevensoverdracht.

Beheer van de DMA-controller

Bij de PC, ISA, EISA en Micro Channel systemen, gebaseerd op de 8237A DMA-controller en afgeleiden hiervan, kan het

19.12 Principes van DMA, Direct Memory Access

gemakkelijk voorkomen dat een programma andere programma's nadelig beïnvloedt wanneer niet de juiste programmeerstappen worden genomen, omdat veel kanaalregisters van de DMA-controller meerdere kanalen beïnvloeden. Daarnaast zijn voor deze computers programma's nodig die rechtstreeks toegang hebben tot de DMA-controller om DMA te kunnen toepassen.
Met andere woorden: het bedrijfssysteem biedt DMA geen ondersteuning. Door dit gebrek aan bescherming moeten programma's die gebruik maken van de DMA-controller bepaalde regels in acht nemen om de integriteit van de DMA-programmering en van de statusinformatie te waarborgen.
Zo worden bijvoorbeeld de 16 bit registers van de 8237A DMA-controller geprogrammeerd via een 8 bit poort, waarbij de DMA-controller beschikt over één interne byte-aanwijzer die bepaalt welk van de twee interne byteposities de gegevens ontvangt van elke 8 bit lees- of schrijfbewerking. Het is noodzakelijk om de byte-aanwijzer vóór het begin van een 16 bit toegangsreeks te wissen, teneinde te kunnen garanderen dat het register correct wordt geladen. Om te voorkomen dat een interrupt-routine de interne byte-aanwijzer verandert, dienen systeeminterrupts te worden geblokkeerd tijdens het wissen van de byte-aanwijzer en het lezen van of schrijven naar het register.

Statusinformatie
Een ander probleem heeft betrekking op de vraag hoe met zekerheid kan worden vastgesteld of een DMA-kanaal gereed is. Het DMA-controller statusregister kan worden gelezen om vast te stellen of één van de kanalen een TC heeft bereikt. Helaas worden door het lezen van het statusregister de TC statusbits voor alle DMA-kanalen in het register gewist. Een programma kan dus niet vertrouwen op de inhoud van het statusregister en moet andere wegen vinden om de status te bepalen van het DMA-kanaal dat wordt gebruikt. Dit houdt ondermeer in het lezen en controleren van de 16 bit adresregisters en telregisters van het kanaal. Voor een nauwkeurige uitlezing moet het kanaal tijdelijk worden afgeschakeld omdat de registerinhouden wijzigen en er twee 8 bit leesopdrachten per register moeten worden uitgevoerd. Dientengevolge wordt het DMA-proces van een specifiek kanaal stopgezet wanneer er statusinformatie wordt opgevraagd. Bij hoge snelheden kan dit net zoveel vertraging opleveren dat er binnenkomende informatie verloren gaat.
Elk programma dat rond de maximale systeemsnelheid gebruik maakt van de DMA-controller dient met dit gegeven rekening te houden.
Een betere methode om de status van het in gebruik zijnde DMA-kanaal vast te stellen, is om het DMA slave-I/O-apparaat de DMA-status te laten bewaken, bijvoorbeeld door gebruik te maken van een op de kaart aanwezige teller of door het opwekken van TC-interrupts. Bij de EISA en Macintosh II DMA-kaart genereert de DMA-controller op het moment van een TC een systeeminterrupt die naar de daarbij behorende interrupt afhandelingsroutine voor het betreffende kanaal springt. In systemen die niet over deze mogelijkheid beschikken, kan de DMA slaaf-I/O-kaart de TC-impuls (die wordt opgewekt door de DMA-controller) bewaken en een interrupt opwekken wanneer er een TC-impuls optreedt die voor hem bestemd is. Met behulp van TC-interrupts wordt het programma snel op de hoogte gesteld dat

de DMA gereed is. Dit is belangrijk in verband met de vele DMA-onderhoudstaken die aan een strikte timing zijn onderworpen. Zo moet bijvoorbeeld een DMA-kanaal opnieuw worden geprogrammeerd en worden herstart vóórdat het slaafapparaat zijn interne buffer laat overlopen. De TC-interrupts kunnen tevens worden gebruikt om bepaalde beperkingen van de DMA-controller te niet te doen, zoals de beperkte lengte van een byte-overdracht (PC, ISA) en het ontbreken van auto-initialisatie (Micro Channel) of buffer chaining (PC, ISA, Micro Channel). De TC-software herprogrammeert simpelweg het kanaal en het lijkt net alsof de DMA-controller zélf over dit vermogen beschikt.

DMA bufferbeheer en data-acquisitie
De eenvoudigste methode van bufferbeheer is enkelvoudig gebufferde gegevensbeheer. Hierbij wordt een vaste hoeveelheid gegevens verzonden naar of overgedragen vanuit een sequentieel geheugenblok. Enkelvoudig gebufferde data acquisitie of opwekking is bruikbaar met een extern getriggerd I/O-apparaat. Het apparaat wacht totdat de trigger is ontvangen alvorens een DMA-aanvraag voor gegevensoverdracht in te dienen.
DMA-controllers voeren deze simpele taak automatisch uit. Een en ander wordt ingeleid door de processor, waarna de afhandeling geheel aan de DMA-controller wordt overgelaten. Wanneer het programma toegang wil hebben tot de gegevens die ze in de buffer zijn geschreven, moet de DMA-controller tijdelijk worden afgeschakeld om de adres- of telregisters te lezen. Het programma kan dan de registerinformatie gebruiken om veilig toegang te verkrijgen tot de buffer. Dit starten en stoppen ondermijnt echter de potentieel hoge snelheid van DMA en moet eigenlijk alleen maar worden gebruikt wanneer DMA toch al zo langzaam is dat het periodiek stoppen van de besturingseenheid geen gegevensverlies oplevert.
In het andere geval moet het programma wachten op een TC-indicatie alvorens toegang te krijgen tot de gegevens in de buffer. Dit betekent dat het programma andere taken kan uitvoeren waarbij onderwijl periodiek de DMA-status wordt getest. Alhoewel de DMA-controller is ontworpen om automatisch enkelvoudig gebufferde data-acquisitie uit te voeren, maken de 64 kB overdrachtslimiet en de 64/128 kB DMA-paginagrens in een PC/AT omgeving het noodzakelijk dat de data acquisitie wordt opgedeeld in meerdere DMA-overdrachtreeksen, waarbij de DMA-controller steeds tussentijds opnieuw moet worden geprogrammeerd.

Verzamelen van pre-trigger gegevens
Bij bepaalde applicaties is het vooraf triggeren (pre-trigger) van de gegevens nodig (gegevens verzamelen voordat de triggerimpuls optreedt). Met behulp van deze techniek kan de besturingssoftware de pre-trigger data acquisitie implementeren door gebruik te maken van de auto-initialisatie eigenschap die is ingebouwd in de meeste DMA-controllers. De DMA-controller plaatst de gegevens cyclisch in de buffer en wanneer het eind van de buffer is bereikt, wordt er opnieuw gestart en plaatst die de gegevens weer aan het begin van de buffer.
Wanneer het I/O-apparaat stopt, leest het besturingsprogramma (driver) de DMA-controller uit om vast te stellen op welke plaats in de buffer het laatste monster werd geschreven. De gegevens in de buffer worden vervolgens opnieuw ingedeeld

19.12 Principes van DMA, Direct Memory Access

zodat het eerste deel van de buffer de pre-trigger gegevens bevat en het tweede deel de post-trigger gegevens.
Omdat voor dit cyclisch verlopende proces auto-initialisatie vereist is, handelt het besturingsprogramma in situaties, waarbij de auto-initialisatie niet in de DMA-controller is ingebouwd, net zoals dat bij de Micro Channel het geval is. Wanneer de buffer een paginagrens overschrijdt of wanneer de lengte de stand van het geprogrammeerde telregister te boven gaat, zal het besturingsprogramma ook gebruik maken van de hierboven beschreven "in stukjes en beetjes" DMA-programmeertechniek.

Dubbel gebufferd DMA-gegevensbeheer
De methode van dubbel gebufferd DMA-gegevensbeheer biedt vele voordelen ten opzichte van het enkelvoudig gebufferde gegevensbeheer.
Terwijl de DMA-controller de ene buffer vult met gegevens, kan een applicatieprogramma de gegevens van een andere buffer lezen zonder rekening te hoeven houden met de geheugenlocatie waartoe de DMA-controller toegang heeft. Deze methode van bufferbeheer staat bekend als "dubbel gebufferd gegevensbeheer" omdat de gegevens voordat ze worden verwerkt vanuit de DMA-buffer moeten worden gekopieerd naar een tweede buffer.
Een circulaire dubbele buffer is nodig wanneer het beschikbare geheugen te klein is voor de hoeveelheid gegevens die moet worden verzameld, of wanneer op elk gewenst moment de meest recente gegevens moeten kunnen worden opgehaald voor het geval de gegevens continu worden verzameld. Dubbel gebufferd gegevensbeheer is met name gunstig wanneer de processor en de DMA-controller met ongeveer dezelfde snelheid werken.
Geen van beiden hoeft dan te wachten tot de ander de toegang tot de buffer heeft afgerond. In theorie zijn twee buffers voldoende. In de praktijk worden er echter meer buffers gebruikt om de tijdverschillen weg te werken ten gevolge van de reactie van de processor en de busvertraging die ontstaat door de tussenkomst van de DMA-controller. Een van de meest populaire toepassingen van dubbel gebufferd gegevensbeheer is het schrijven van data naar schijfheugenbestanden tijdens de data-acquisitie. Deze methode is ook handig bij het ontwerpen van een interactieve applicatie die een real-time afbeelding actualiseert, bijvoorbeeld een op schaal gebrachte of een uit de verzamelde gegevens afgeleide afbeelding.

Bufferbeheer in virtuele geheugensystemen
Het DMA-bufferbeheer wordt ingewikkelder wanneer de processor gebruik maakt van een virtueel geheugenbeheersysteem of op een andere manier informatie dynamisch in het geheugen verplaatst. Het Macintosh II systeem beheert het stapelgeheugen van de applicatie zodanig dat de gegevens in het geheugen op elk moment kunnen worden verplaatst. Dit vereist dat applicaties bij het toekennen van geheugenruimte aan het besturingssysteem om een onverplaatsbaar geheugen vragen voor gebruik met DMA.
Bij de Macintosh II computers zijn ook voldoende gebieden met virtueel geheugen geïmplementeerd. Hierbij verschillen de door de processor gebruikte adressen (logische adressen) van de adressen die worden gebruikt door de NuBus DMA-controller (fysische adressen).
Dit komt omdat apparaten op de NuBus voor het virtuele geheugensysteem extern zijn.

19.12 Principes van DMA, Direct Memory Access

Type	DMA controller	Driver	Applicatie
Enkelvoudig gebufferd (natriggering)	buffer trigger DMAC vult de buffer eenmaal nadat de trigger is ontvangen.	- Start de DMA controller (DMAC). - Rapporteert de acquisitie status.	- Start de acquisitie. -Wacht tot de acquisitie gereed is alvorens de buffer te verwerken.
Enkelvoudig gebufferd (voortriggering)	buffer trigger DMAC maakt gebruik van auto-initialisatie om de buffer circulair te vullen totdat de trigger wordt ontvangen.	- Start de DMAC* -Rapporteert de acquisitie status. -Geeft de buffer vrij - als DMA klaar is.	- Start de acquisitie. -Wacht tot de acquisitie gereed is alvorens de buffer te verwerken.
Dubbel gebufferd (alleen natriggering)	block buffer DMAC maakt gebruik van chaining om elk blok te vullen totdat het totale aantal blokken is verzameld of het proces door de applicatie wordt gestopt.	- Start de DMAC. -Kopieert de aangevraagde blokken naar de applicatiebuffer zodra ze beschikbaar komen. -Rapporteert de acquisitie status.	- Start de acquisitie. -Haalt de gegevens in blokken op; kan zowel de oudste als de meest recente gegevensblokken ophalen. - Stopt de acquisitie wanneer er voldoende gegevensblokken zijn verzameld.
Frame georiënteerd (voor- en natriggering)	frame buffer Net als bij het enkelvoudig gebufferde gegevensbeheer (natriggering of voortriggering) verwerkt DMAC elke trigger en elk frame. Het proces stopt wanneer het totale aantal frames is verzameld of wanneer de applicatie het proces beëindigt.	-Start de DMAC opnieuw wanneer elk frame is verzameld. - Kopieert de aangevraagde frames naar de applicatiebuffer zodra ze beschikbaar komen - zonodig nadat ze zijn vrijgegeven. -Rapporteert de acquisitie status.	- Start de acquisitie. -Haalt n frames met gegevens op; kan zowel de oudste als de meest recente frames ophalen. - Stopt de acquisitie wanneer er voldoende frames zijn verzameld.

Figuur 3/19.12-8: Overzicht van DMA-bufferbeheer.

Daarnaast kan buffer chaining worden toegepast om gegevensbuffers te verwerken die fysisch niet aan elkaar grenzen maar die volgens de logische adressen wel aangrenzend zijn.

Samenvatting
In de tabel van figuur 3/19.12-8 is een overzicht gegeven van de verschillende systemen die zijn ontwikkeld voor het beheren van de DMA-buffer.

Conclusie

Overhead
Bij het programmeren van een DMA-controller moet rekening worden gehouden met enige overhead. Deze overhead moet over de totale hoeveelheid overgedragen data worden omgeslagen. Wanneer de over te dragen hoeveelheid gegevens gering is, kan de processor de overdracht wellicht sneller uitvoeren. Dit "break-even" punt is ook afhankelijk van de in het DMA-slaafapparaat aanwezige (buffer)geheugenomvang.
Een voorbeeld: wanneer de slaafbuffer N locaties heeft en er N of minder gegevens moeten worden overgedragen, is het raadzaam om de slaafbuffer te vullen. Hierna stuurt de slaaf dan een interrupt naar de processor zodat (nadat alle gegevens zijn verzameld) de processor deze op topsnelheid kan inlezen. Wordt gebruik gemaakt van DMA dan moet de overhead worden berekend en het programma zodanig worden opgezet dat de processor de overdrachten uitvoert wanneer de hoeveelheid gegevens gering is en de processor niet hoeft te wachten tussen de gegevensoverdrachten. In applicaties zoals "timed data acquisitie" komt doorgaans geen overhead voor omdat de acquisitie vaak wordt gestart door een trigger. De DMA-controller wordt dan ingesteld voordat er een trigger wordt ontvangen.
De overhead ten gevolge van het instellen van de DMA-controller kan worden gezien als offset wanneer de DMA-controller de gegevens sneller kan overdragen dan de processor, wanneer de applicatie de snelheid vereist, of wanneer de applicatie tegelijk met de gegevensoverdracht andere taken moet verrichten.

Voor- en nadelen van DMA
De DMA hardware en software technologieën worden zowel gebruikt om hoge doorvoersnelheden te bereiken alsook om de systeembenutting te vergroten. Deze doelstellingen kunnen worden bereikt wanneer voor de gegevensoverdracht een achtergrondmechanisme wordt toegepast waardoor de processor wordt ontlast. De gebruikers van data-acquisitie systemen zijn zich zeer wel bewust van de grote voordelen van deze "data-acquisitie op de achtergrond" en DMA-oplossingen zijn zeer populair geworden. Hoewel ten gevolge van DMA de systeembenutting toeneemt en er een hoge doorvoersnelheid kan worden bereikt, moet er ook rekening mee worden gehouden dat interrupts soms toch bepaalde voordelen bieden ten opzichte van DMA. Een DMA-controller kan alleen gegevens van de ene locatie verplaatsen naar de andere. Het is echter vaak noodzakelijk om gegevens "on the fly" te verwerken, bijvoorbeeld zoeken naar een triggervoorwaarde in de gegevens alvorens deze in het geheugen op te slaan, of het implementeren van een zeer snelle procesbesturingslus. Interruptgestuurde acquisitie is voor deze applicaties noodzakelijk omdat er op individuele aspecten bewerkin-

gen moeten worden uitgevoerd. Daarnaast zijn er meestal minder DMA-kanalen beschikbaar dan interruptkanalen. De meeste DMA-software maakt afwisselend gebruik van DMA en interrupts zodat de beschikbaarheid van kanalen transparant is voor het applicatieprogramma. Een ander voordeel is dat interruptroutines gegevens van een aantal locaties kunnen overdragen naar verschillende gegevensbuffers, terwijl er voor elke locatie en buffer een DMA-kanaal nodig zou zijn. De flexibiliteit van interrupts moet worden afgewogen tegen een intensiever processor-gebruik, een snellere dienstverlening en, in bepaalde gevallen, een hogere doorvoersnelheid bij DMA.

DMA en lage gegevenssnelheid

Ook wanneer de gegevens langzaam binnenkomen kan DMA worden toegepast, hoewel de voordelen van een betere benutting van de processor en de snellere dienstverlening dan niet tot hun recht komen.

In feite levert het bij lage snelheden wachten tot er een blok van een buffer is gevuld geen enkel voordeel op. In deze gevallen moet de DMA-controller worden gelezen om de status te bepalen, hetgeen een sneller kanaal kan storen. Om te kunnen bepalen of het zinvol is om DMA voor de langzamere kanalen toe te passen, dient de totale DMA-systeembenutting bekend te zijn.

3/19.13

Principes van video-integratie

Inleiding

Groeimarkt video

Video op het scherm van de PC-monitor is een van de grootste groeimarkten van dit moment. Het algemene principe is dat in een klein venstertje op het scherm van de monitor bewegende videobeelden worden vertoond, waarvan de gegevens afkomstig zijn van externe of interne bronnen: videocamera, videorecorder, TV-tuner, CD-ROM, DVD, Video-CD, beeltelefoon. De bewegende videobeelden worden dus op de een of andere manier geïntegreerd in de overige vrij statische informatie op het beeldscherm.
Als voornaamste toepassingen kunnen vermeld worden:
- video conferencing;
- opleidingen;
- ontspanning;
- eigen videoproducties;
- TV on PC.

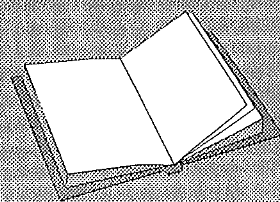
Video conferencing

Het vergaderen op afstand via de PC wordt algemeen beschouwd als een van de voornaamste toekomstige toepassingen van zakelijk gebruik van video op de PC. Dank zij breedbandige netwerken zoals ISDN en satellietinternet en moderne compressietechnieken is het reeds nu mogelijk het videosignaal van een videocamera onder digitale vorm real time van de ene PC naar de andere te versturen. De eenmalige investering die noodzakelijk is voor videovergaderen zal helpen een deel van de jaarlijkse reiskosten uit te sparen. Een tweede voordeel is een niet onaanzienlijke besparing op vergader- en reistijden.

Opleidingen

Interne opleidingen zullen steeds vaker verzorgd worden onder de vorm van multimedia-presentaties op de monitor van de PC. Statische tekst, illustraties en grafieken kunnen aangevuld worden met gesproken tekst, geluidseffecten, animaties en echte video. De bron is dan meestal een CD-ROM, die voldoende capaciteit heeft om tientallen minuten video-informatie te bevatten. In de meeste gevallen zal voor dit soort toepassingen video van lage kwaliteit volstaan, zoals reeds nu wordt verkregen onder de vorm van soft-

LEES OOK:
Hoofdstuk 3/19.9
Hoofdstuk 5/11.1
Hoofdstuk 5/17.1
Hoofdstuk 5/17.2
Hoofdstuk 5/20.1



warematige CODEC's. Bekende formaten zijn Indeo, Cinepack MPEG en AVI.

Ontspanning
De belangrijkste groeimarkt in het thuissegment is ongetwijfeld de DVD en alles dat hierbij hoort. Dank zij deze standaard kan men ongeveer drie uur audio- en videogegevens van hoge kwaliteit op één CD-ROM onderbrengen. Nu zullen er maar weinig mensen zijn die naar films willen kijken via een klein venstertje op het scherm van de monitor. De techniek biedt echter wel heel veel mogelijkheden in de spelletjessector.
Computergames in de actiesector zullen zich afspelen tegen realistische achtergronden, die worden gegenereerd via DVD-gecodeerde video.

Eigen videoproducties
In dit geval spreekt men van "life video". Het videosignaal kan afkomstig zijn van een camcorder of van een videorecorder. Geëigende software laat toe diverse videobronnen met elkaar te mengen en videobeelden te voorzien van titels en/of ondertitels. De aldus gemonteerde productie kan weer als analoog videosignaal weggeschreven worden naar een videorecorder. Binnen niet al te lange tijd zullen analoge videosignalen via betaalbare MPEG-codeerkaarten als digitale datastroom naar CD-R's weggeschreven kunnen worden.
Dan ontstaan mogelijkheden om vakantie-, instructie-, demonstratie- en verkooppromotiemateriaal onder de vorm van een multimedia presentatie in eigen beheer te ontwerpen en te produceren.

TV on PC
Tot slot bestaat de mogelijkheid TV-signalen op te vangen via speciale kaarten en de uitzendingen in een venstertje op de monitor te vertonen. Ook hiervan wordt verwacht dat dit een segment is met zeer groot groeipotentieel. TV-beelden kunnen "bevroren" worden en omgezet in de bekende grafische formaten TIF, JPEG of BMP en rechtstreeks verwerkt worden in grafische programma's.

Hardware
Het be- en verwerken van video-informatie vraagt nogal wat rekencapaciteit en het belasten van de systeemprocessor hiermee zou tot grote systeemvertragingen leiden. Vandaar dat voor al deze toepassingen speciale hardware wordt geleverd. Speciale videoprocessoren nemen het rekenintensieve werk over van de systeemprocessor en zorgen ervoor dat de videogegevens worden ingevoegd in de gegevens die de bestaande grafische kaart naar de monitor stuurt.
Voor dat invoegen van de videogegevens in de datastroom zijn verschillende systemen ontwikkeld, die in dit hoofdstuk aan de orde komen.

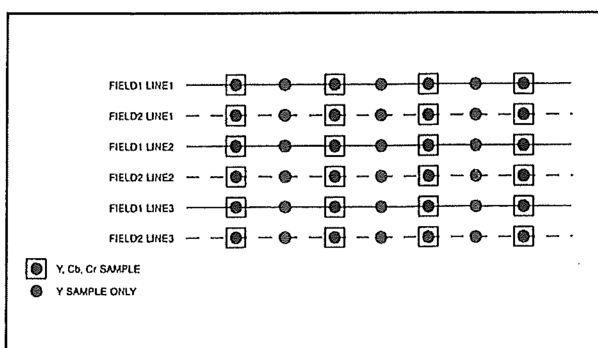
Het verwerken van video

Inleiding
Zoals reeds geschreven vraagt het verwerken van video in een PC nogal wat rekencapaciteit.
De eerste vraag die beantwoord moet worden is waarom dit zo is. In wezen zijn er vier factoren die aan de orde komen bij het integreren van video op het scherm van de monitor:
- kleurconvertering;
- resolutie herstel;
- schaling;
- menging.

19.13 Principes van video-integratie

Kleurconvertering

Alle bestaande videosystemen maken gebruik van een kleurschema dat bekend staat onder de naam "YUV", ook wel "Y, Cb, Cr" genoemd. De kleurinformatie van het beeld wordt hierbij niet gedefinieerd door de drie basiskleuren rood R, groen G en blauw B, zoals bij een PC-monitor, maar door de helderheid Y en twee kleurcomponenten U en V (Cb, Cr). Het voordeel van het YUV-formaat is dat er slechts 16 bits noodzakelijk zijn voor echte "true color", terwijl het RGB-systeem hiervoor 24 bit nodig heeft. Een tweede voordeel van het YUV-formaat is dat men een aanmerkelijke datareductie kan bereiken door handig gebruik te maken van de beperkingen van het menselijke oog. Het blijkt namelijk dat het oog zeer gevoelig is voor helderheidsverschillen, maar tamelijk ongevoelig voor kleurverschillen. Het is dus niet noodzakelijk van ieder beeldpunt zowel Y- als U- en V-waarden te versturen. Aan de hand hiervan zijn verschillende compressietechnieken ontwikkeld, zoals YUV-4:2:2, YUV-4:1:1 en YUV-4:2:0.



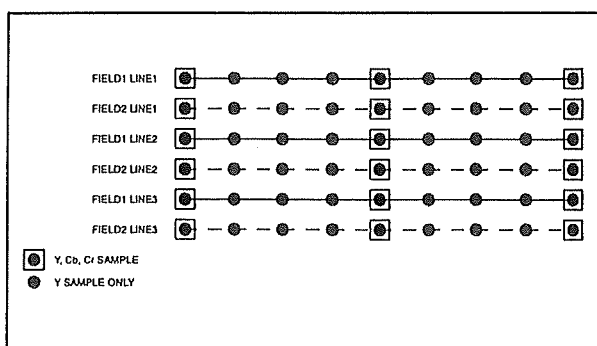
Figuur 3/19.13-1: Het codeerschema van YUV-4:2:2.

YUV-4:2:2

Bij YUV-4:2:2 wordt binnen één beeldlijn maar van ieder tweede pixel de kleurinformatie gecodeerd. Dit schema is voorgesteld in figuur 3/19.13-1. Men start men het berekenen van de Y-, U- en V-waarden van het eerste beeldpunt van iedere lijn. Van het tweede beeldpunt wordt alleen de Y-waarde berekend. Nadien worden weer de drie waarden vastgesteld. Dit schema levert een datareductie van een/derde op.

YUV-4:1:1

Bij YUV-4:1:1, voorgesteld in figuur 3/19.13-2, wordt slechts van ieder vierde pixel de kleurwaarden berekend en in de datastroom opgenomen. Omdat per pixel één Y-waarde en twee U/V-waarden noodzakelijk zijn, heeft het YUV-4:1:1 schema een datareductie van 50 % tot gevolg.



Figuur 3/19.13-2: Het YUV-4:1:1 schema.

YUV-4:2:0

Bij YUV-4:2:0 wordt van ieder beeldpunt de Y-waarde berekend, maar wordt voor de U- en V-waarden het gemiddelde berekend van vier naast elkaar liggende pixels van opeenvolgende beeldlijnen. Dit schema, voorgesteld in figuur 3/19.13-3, wordt bijvoorbeeld gebruikt bij het digitaliseren van videobeelden volgens de MPEG-1 standaard en de H-261 standaard.

Van YUV naar RGB

Omdat monitoren van PC-systemen altijd werken volgens het RGB-schema, moeten de inkomende YUV-gegevens in real time

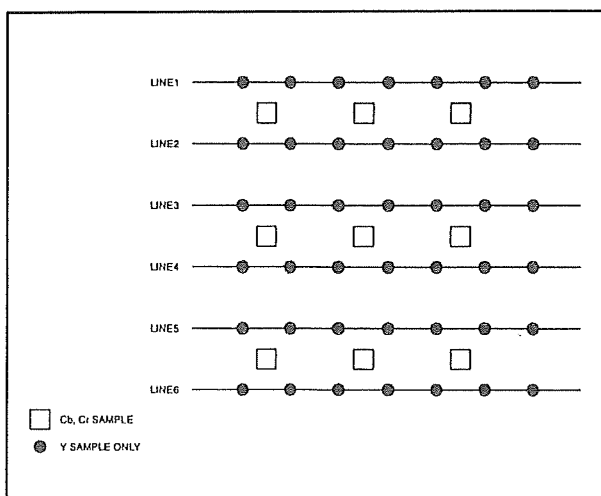
omgerekend worden naar RGB-waarden. Hiervoor bestaan verschillende omrekenalgorithmen, maar het meest gebruikte zijn de onderstaande formules, die gestandaardiseerd zijn in de CCIR-aanbeveling 601. De CCIR is het "International Radio Consultative Committee", een internationale commissie die zich bezig houdt met het opstellen van internationale standaarden op de gebieden audio en video. Volgens CCIR-601 is:

G = Y - 0,698 V - 0,338 U + 116,56

R = Y + 1,371 V - 191,45

B = Y + 1,732 U - 237,75

Dat zijn eenvoudige formules, zeker voor een processor, maar deze berekeningen moeten wél 25 keer per seconde voor ieder van de tienduizenden beeldpunten van een beeld herhaald worden!



Figuur 3/19.13-3: Het YUV-4:2:0 schema, onder andere gebruikt bij MPEG-1.

Herstellen van de resolutie

Alvorens het systeem echter de RGB-waarden van ieder punt berekent, moeten de gereduceerde schema's omgezet worden in YUV-waarden voor ieder beeldpunt. In de meeste gevallen worden gemiddelde waarden voor de ontbrekende U- en V-waarden berekend, maar ook deze wiskundige bewerkingen moeten 25 keer per seconde voor iedere beeldpunt worden uitgevoerd.

Schaling

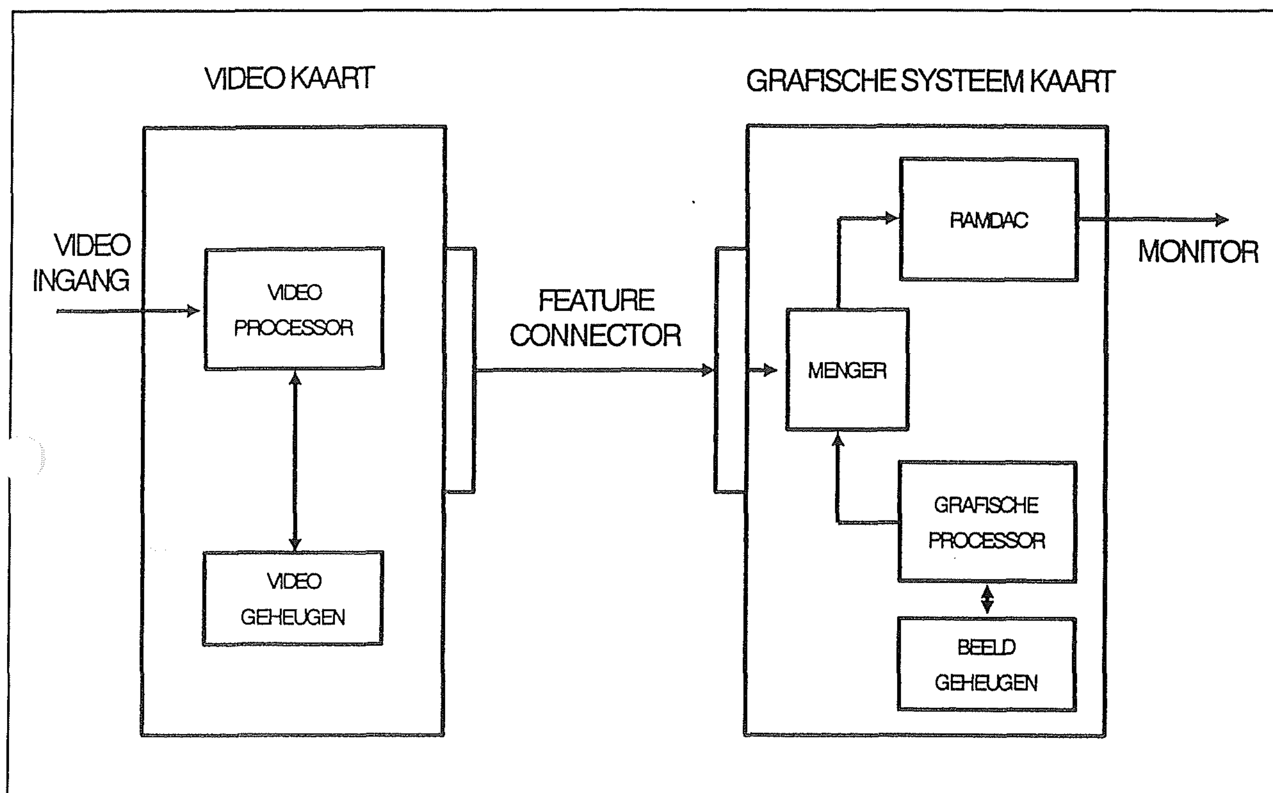
Een derde rekenintensieve bewerking is het schalen van de videogegevens naar het juiste formaat. Het zal immers maar zelden voorkomen dat het videovenstertje op het scherm net zoveel horizontale en vertikale beeldpunten bevat als het gedigitaliseerde videobeeld. Dus moet er geïnter- of geëxtrapoleerd worden, een procédé waar heel wat wiskunde aan te pas komt.

Het simpelweg weglaten van beeldpunten (formaat verkleinen) of dupliceren van beeldpunten (formaat vergroten) leidt tot aanzienlijk verlies van beeldinformatie of tot een blokvormig beeld. Dat laatste verschijnsel is iedereen die wel eens een AVI-venstertje tot volle-scherm afmetingen heeft vergroot welbekend! Vandaar dat ingewikkelde wiskundige interpolatietechnieken worden toegepast, waarbij nieuwe beeldpunten worden berekend uit de informatie van nabuurpunten. De kwaliteit van het beeld in het videovenstertje is voornamelijk afhankelijk van de manier waarop deze interpolatie wordt uitgevoerd.

In de horizontale richting gaat dit interpoleren vrij eenvoudig, omdat de gegevens van de naast elkaar liggende beeldpunten in seriële vorm worden aangeleverd en een zeer kleine geheugencapaciteit noodzakelijk is om de gegevens gedurende de berekeningen op te slaan.

Bij het interpoleren in verticale richting moet de processor beschikken over alle gegevens van minstens twee beeldlijnen, iets dat veel meer geheugen kost.

19.13 Principes van video-integratie



Figuur 3/19.13-4: Het principe van video overlay.

Menging

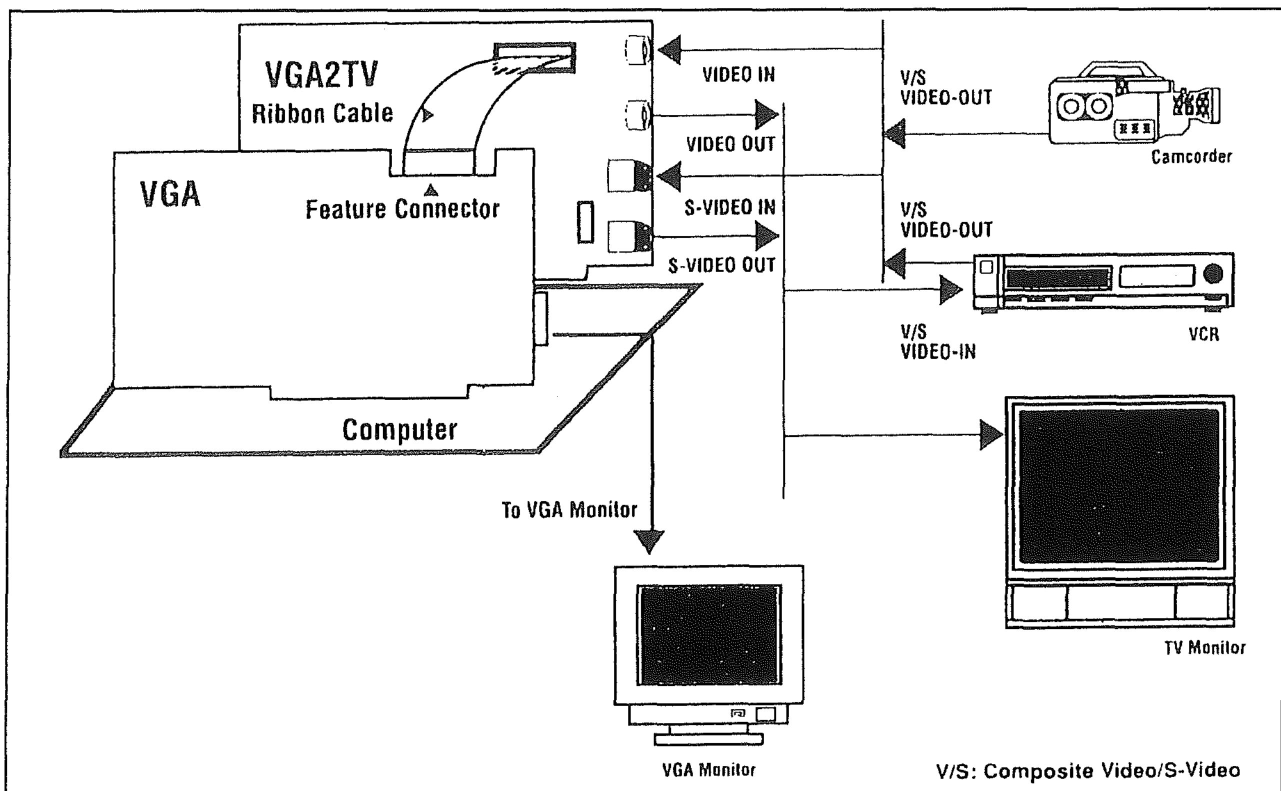
Een laatste tamelijk rekenintensieve activiteit is het mengen van de videogegevens met andere gegevens op het beeldscherm. Als het actieve videovenster over andere vensters heen ligt en men maakt een van deze andere vensters actief, dan zal een deel van het videovenster overschreven worden door de inhoud van het geactiveerde venster. Als men nadien het videovenster weer activeert zal de informatie in het sluitende venster weer naar de "achtergrond" moeten verdwijnen. Hetzelfde probleem ontstaat als men een menu opent in een videovenster. Ook dan moet de momentele video-informatie gedeeltelijk vervangen worden door andere gegevens. Hoewel dit een probleem is dat zich met alle vensters voordoet is het grote verschil dat de meeste vensters vrij statische informatie bevatten, die niet 25 keer per seconde hernieuwd wordt. Met een videovenster is dat wél het geval, hetgeen veel grotere eisen stelt aan de rekenkracht van de processor die zorgt voor de beeldopbouw.

Principes voor video-integratie

Inleiding

Het integreren van "life" video op het beeldscherm vergt dus nogal wat inspanningen van de processor. Vandaar dat tegenwoordig vrijwel steeds wordt gewerkt met specifieke videoprocessoren die geheel of gedeeltelijk zorgen voor het verwerken van de videogegevens en deze naar het beeldscherm sturen.

19.13 Principes van video-integratie



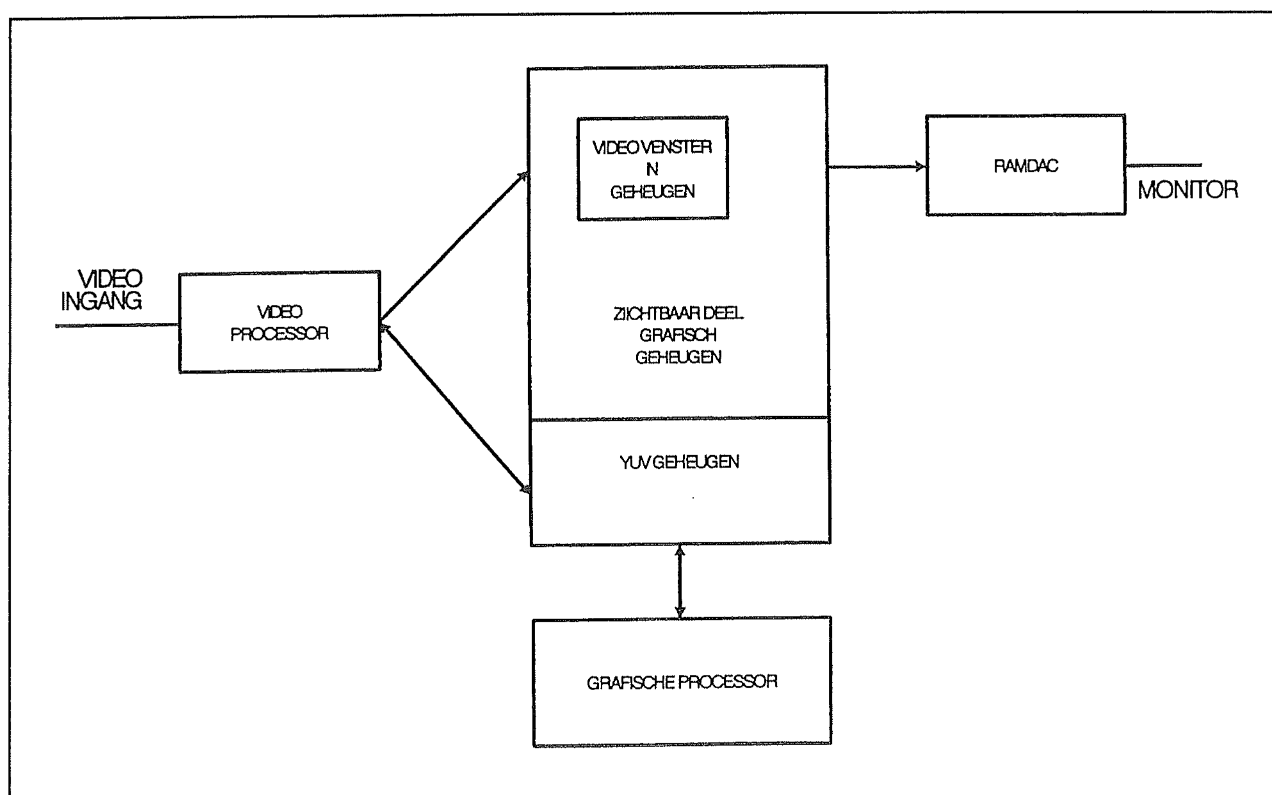
Figuur 3/19.13-5: Een video kaart kan via de feature connector met de standaard grafische kaart van het systeem verbonden worden.

Hierbij doet zich echter een groot probleem voor. In ieder PC-systeem is een videokaart aanwezig, die via een kabel verbonden is met de monitor. Dit is een gesloten systeem en het is niet zo eenvoudig om in dit systeem "in te breken". Toch is dat noodzakelijk als men de gegevens die door een speciale videoprocessor worden aangeleverd naar het beeldscherm wil sturen.
Hiervoor zijn in de loop der jaren vier systemen ontwikkeld, namelijk:
- video overlay techniek;
- shared frame buffer architecture;
- videoport architecture;
- backend scaling technology.

Video overlay techniek

Bij deze techniek wordt gebruik gemaakt van de "VGA Feature Connector" die op vrijwel iedere grafische kaart aanwezig is. Het principe van video overlay is geschetst in figuur 3/19.13-4.
De videogegevens worden verwerkt op een eigen kaart die over een eigen geheugen beschikt. Via de kabel die de twee feature connectoren verbindt, worden de RGB-gegevens van de videokaart naar een menger op de grafische kaart van het systeem gevoerd. Hier worden deze gegevens gemengd met de RGB-gegevens die door de grafische kaart zélf worden aangemaakt. Het gemengde signaal stuurt de RAMDAC aan, die de RGB-gegevens omzet in drie analoge signalen waarmee de monitor wordt aangestuurd. Het gebruik van de feature connector lijkt op het eerste zicht veel voordelen te hebben. Zo kan men de bestaande grafische kaart steeds uitbreiden met nieuwe hardware.

19.13 Principes van video-integratie



Figuur 3/19.13-6: Het principe van shared frame buffer architecture.

De enige voorwaarde is dat de uitbreidingskaart over een feature connector beschikt. Een typische opstelling van grafische kaart plus video kaart is weergegeven in figuur 3/19.13-5. Diverse fabrikanten brengen bijvoorbeeld MPEG-kaarten op de markt die op deze manier in een systeem geïntegreerd kunnen worden.
Toch kleven er behoorlijk wat bezwaren aan het overlay principe. De standaard VGA feature connector werkt slechts met een databus van 8 bit en de gegevens worden dus vrij traag overgebracht. Bovendien is het systeem alleen geschikt voor de standaard VGA-resolutie van maximaal 640 bij 480 pixels en een kleurendiepte van 8 bit. Dit komt overeen met maximaal 256 kleuren.

VAFC

De beperkingen van de VGA feature connector kunnen opgeheven worden door gebruik te maken van de nieuwe standaard VAFC, oftewel "VESA Advanced Feature Connector".
Toch kleven er aan het idee van doorlussing van twee grafische kaarten andere nadelen. Er zijn steeds twee geheugens noodzakelijk en de softwarematige integratie van en aanpassing tussen de twee kaarten wil wel eens heel moeizaam verlopen. Vandaar dat steeds vaker wordt gekozen voor oplossingen, waarbij videofuncties een geïntegreerd deel uitmaken van de standaard grafische kaart.

19.13 Principes van video-integratie

Shared frame buffer architecture

Een van de eerste oplossingen hiervoor is de "shared frame buffer architecture". Hierbij wordt gebruik gemaakt van één beeldgeheugen. De videogegevens worden eerst onder de vorm van YUV-data in een niet op het beeldscherm afgebeelde deel van het geheugen opgeborgen. De gegevens kunnen hieruit snel opgeroepen worden, omgezet naar RGB en nadien in het zichtbare bereik van het grafisch geheugen opgeborgen worden.

Het blokschema van een grafische kaart die volgens dit principe werkt is getekend in figuur 3/19.13-6.

Nadeel van deze architectuur is dat er talrijke lees- en schrijfacties naar het grafisch geheugen noodzakelijk zijn. Om een en ander met voldoende snelheid te kunnen realiseren moeten er hoge eisen gesteld worden aan de organisatie van en de toegangstijd tot het geheugen. Vandaar dat er gewerkt wordt met een 64 bit brede interne bus en vaak snelle EDO-RAM's worden toegepast. Een en ander heeft tot gevolg dat overdrachtsnelheden van 500 MB/s mogelijk zijn.

In de modernste kaarten wordt gebruik gemaakt van VRAM's, waardoor het tijdrovende verversen van de geheugeninhoud niet noodzakelijk is. Voor het schalen van het videobeeld in het videovenster wordt in de moderne systemen gebruik gemaakt van bilineaire interpolatie. De beeldpixels van het videovenster worden hierbij allemaal berekend uit de nabijgelegen pixels van de video-informatie.

Nadeel

Een nadeel van de shared frame buffer technologie is dat de kleurendiepte van het videovenster identiek is aan de kleurendiepte van het volledige beeld. Omdat de gegevens van de grafische processor en de video processor achter elkaar naar de RAMDAC worden gestuurd, moeten deze gegevens hetzelfde formaat hebben. Werkt men dus met slechts 256 kleuren, dan zal men ook in het videovenster niet meer kleuren kunnen aanschouwen! Een en ander heeft tot gevolg dat men voor true color video met grafische geheugens van 4 tot 8 MB moet werken, ook als men de daarbij horende hoge resolutie en true color over het gehele beeldscherm niet per se nodig heeft.

Dithering

Om dit probleem te omzeilen wordt soms "dithering" in het videovenster toegepast. Hierbij wordt uitgegaan van slechts 256 kleuren. Door de video beeldpunten te vergroten en ieder beeldpunt in het videovenster samen te stellen uit verschillende pixels, die ieder een van de 256 kleuren kunnen hebben, kan men veel meer mengkleuren genereren. Het zal duidelijk zijn dat dithering natuurlijk een desastreuze invloed heeft op de resolutie van het videobeeld! Vandaar dat deze techniek niet aangeraden kan worden voor professionele toepassingen en men minstens met 16 bit kleur voor het gehele scherm moet werken om kwalitatief hoogwaardige video in een venster te verwachten.

Videoport architecture

Een andere architectuur wordt onder andere door Tseng Labs toegepast. Bij deze zogenoemde "Videoport"-architecture wordt gewerkt met een videoprocessor die de YUV-gegevens rechtstreeks omzet in RGB-data. Deze worden in real time overgedragen naar de eigenlijke grafische processor via een zogenoemde Video-port. De grafische controller synchroniseert de videocontroller, zodat de videogegevens

op het juiste moment worden aangeleverd om deze rechtstreeks in de juiste adressen van het grafisch geheugen op te slaan. Deze technologie heeft als voordeel dat er geen extra geheugen nodig is en dat er heel wat minder heen en weer geschreven en gelezen moet worden. Ook nu blijft men echter met het nadeel zitten dat de kleuromvang van de videogegevens identiek is aan deze die voor het gehele beeld werd ingesteld. Wil men video van behoorlijke kwaliteit, dan zal de grafische kaart voldoende geheugen moeten bevatten om het gehele scherm met minstens 16 bit diepe data te beschrijven.

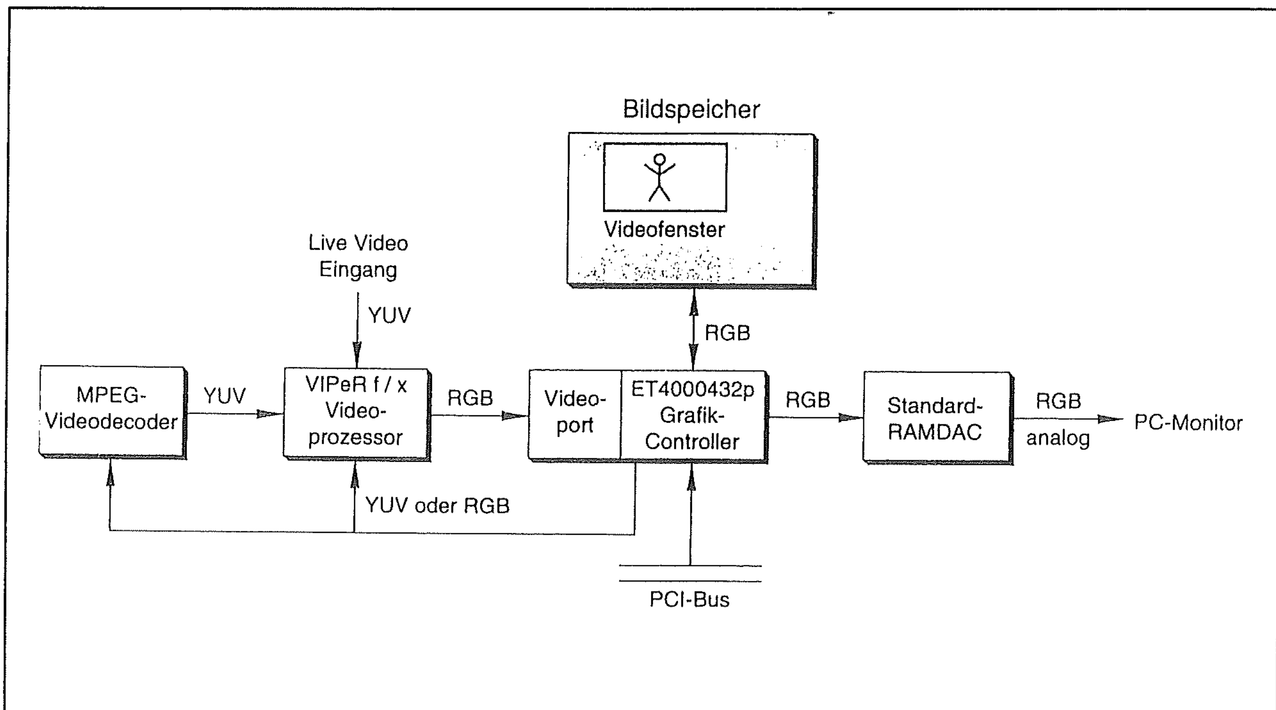
Alles in een

Specifieke eigenschap van zowel de shared frame buffer als de videoport architectuur is dat de grafische processor en de speciale videoprocessor op een en dezelfde kaart zitten en een vrijwel gesloten systeem vormen. Het is niet mogelijk het systeem uit te breiden met een extra kaart. Vandaar dat aan de toegepaste videoprocessor de eis wordt gesteld dat deze in staat is diverse soorten video-informatie te verwerken.

De door Tseng Labs ontwikkelde videoprocessor VIPeR f/x kan diverse YUV- en RGB-formaten verwerken tot een maximale resolutie van 1.280 bij 580 pixels en een kleurendiepte van 24 bit. De processor heeft een afzonderlijke ingang waarmee een MPEG-1 datastroom kan aangevoerd worden. Hierdoor is het systeem in staat alle tot nu toe gebruikelijke videostandaarden te verwerken:

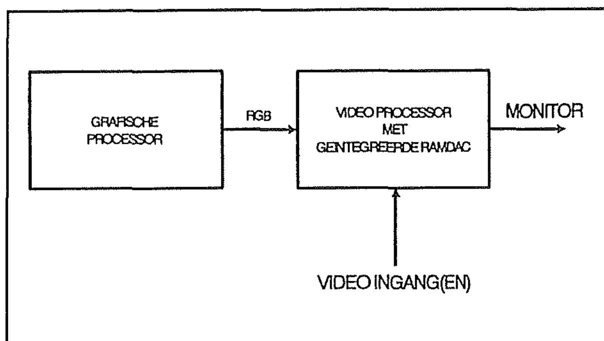
- real time video van een camera of videorecorder;
- MPEG-1 gecodeerde datastromen van een CD-ROM of Video-CD;
- de standaard CODEC's zoals AVI.

Het blokschema van een systeem met video-port architectuur is getekend in figuur 3/19.13-7.



Figuur 3/19.13-7: Het blokschema van een systeem opgebouwd volgens de video-port architectuur.

19.13 Principes van video-integratie



Figuur 3/19.13-8: Het principe van backend scaling.

Backend scaling technology

Een groot bezwaar van de tot nu toe behandelde architecturen is dat de kleurendiepte van de video-informatie gelijk is aan deze van de rest van het beeldscherm. Maar weinig PC-gebruikers zullen de behoefte hebben om de kleurendiepte van het gehele beeldscherm in te stellen op 24 bit. Bij de meeste toepassingen heeft het gebruik van true color geen enkel praktisch nut en levert in feite alleen nadelen op. Immers, er moeten veel meer data verwerkt worden, waardoor de snelheid van de beeldopbouw merkbaar afneemt. Ideaal zou zijn als men er in slaagde een architectuur te ontwerpen waarbij men, bij wijze van spreken, voor ieder venster de noodzakelijke kleurendiepte afzonderlijk kon instellen. Een architectuur die dit mogelijk maakt is de "backend scaling technology".

Bij deze technologie wordt de verschillende datastromen eerst op het laatste van het verwerkingsproces met elkaar gemengd. Vandaar de benaming "backend". Het vereenvoudigde blokschema van een dergelijk systeem is getekend in figuur 3/19.13-8. Eigenschap van deze systemen is dat de normale beeldgegevens en de videogegevens in afzonderlijke delen van het geheugen worden opgeslagen. De twee datastromen kunnen dus ieder in hun eigen formaat in het geheugen worden opgeslagen, omdat de inhoud van het geheugen nu geen pixel voor pixel representatie van het beeldscherm hoeft te zijn. Na de noodzakelijke bewerkingen, dus:

- herstellen van de resolutie;
- omzetten naar RGB;
- schaling;

worden de videogegevens en de normale beeldgegevens op hun weg naar de RAMDAC met elkaar "gemengd". De beeldprocessor zorgt ervoor dat de RAMDAC de twee datastromen onafhankelijk van elkaar verwerkt en omzet in analoge signalen voor het sturen van de monitor.

Dank zij dit procédé is het mogelijk een klein true color videovenster met een kleurendiepte van 24 bit weer te geven op een scherm dat is ingesteld op 256 kleuren, dus 8 bit. Het zal duidelijk zijn dat deze technologie niet alleen veel beeldgeheugen spaart, maar de snelheid van de beeldopbouw ten goede komt.

Moderne video processoren

Diverse fabrikanten brengen op dit moment video processoren op de markt, die werken volgens het principe van de backend scaling.

- S3
  S3 heeft de Trio64V+ in zijn programma. Deze chip maakt gebruik van een eigen interne interface, die "Scenic Highway" wordt genoemd. De chip kan YUV- en MPEG-datastromen verwerken en kan standaard DRAM-geheugen aansturen. De schakeling heeft een geïntegreerde RAMDAC met een bandbreedte van 135 MHz.
- Alliance Semiconductor
  De ProMotion 3210 van deze firma biedt ongeveer dezelfde functionaliteit als de Trio64V+, alleen ontbreekt een YUV-ingang.

19.13 Principes van video-integratie

- Cirrus Logic
  De CL-GD5440 van Cirrus Logic werkt met een ingebouwde RAMDAC met een snelheid van 135 MHz en heeft een geheugenbus met een breedte van slechts 32 bit voor het aansturen van standaard DRAM's. Deze chip heeft een ingang, waarop "live video" kan worden aangesloten.
- Trident
  De TGUI9680 van Trident werkt eveneens met een RAMDAC van 135 MHz en heeft een 64 bit brede databus waarop behalve normale DRAM's ook EDO-RAM's en WRAM's kunnen worden aangesloten. Dit laatste type geheugen werkt met een soort van EDO-architectuur. De TGUI9680 kan de controle over het geheugen overdragen aan perifere schakelingen, zoals de speciale MPEG-processor TMVP 9550.

19.13 Principes van video-integratie

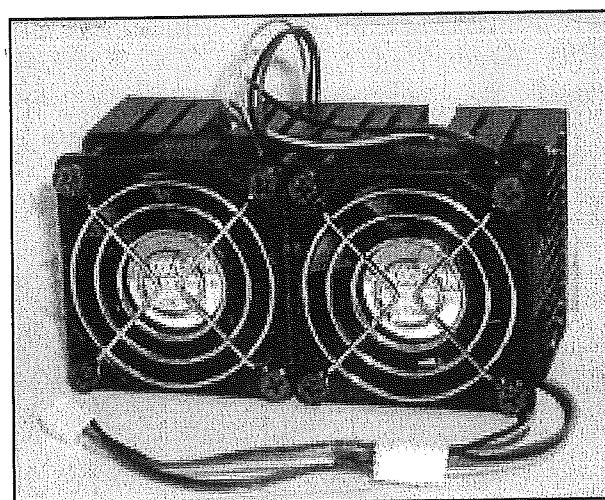
3/19.14

Waterkoelingstechnologie voor snelle processoren

Hitte wordt een steeds groter probleem

Een échte "radiateur"

Wie wel eens de motorkap van een PC openschroeft stelt vast dat de elektronica steeds minder plaats inneemt. De kast is zo goed als leeg op één onderdeel na dat zéér prominent aanwezig is: de "radiateur" op de processor. Er wordt met opzet van een radiateur en niet van koelplaat gesproken, zoals vroeger het geval was. Het onderdeel vertoont inderdaad wel wat gelijkenis met de radiateur van een auto. Een groot massief aluminium meestal zwart geschilderd blok, voorzien van één of soms wel twee ventilatoren voor geforceerde luchtkoeling, zie figuur 3/19.14-1.



Figuur 3/19.14-1: De koeling van een moderne snelle processor vertoont meer gelijkenis met een radiateur dan met een ouderwetse koelplaat.

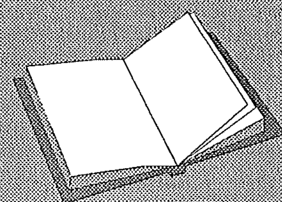
Stroom wekt warmte op

Elektrische stroom die door een geleider vloeit wekt warmte op. Een natuurwet waar zelfs moderne PC-bouwers niet aan kunnen ontkomen. En is een moderne GHz-processor, met zijn miljoenen onderdelen en zijn kloksnelheid van meer dan 1 GHz, vloeit nogal wat stroom. De processor wordt dus flink heet, zeker als men voortdurend processorintensieve toepassingen draait, zoals het beluisteren van MP3-muziek of het afspelen van DVD's.

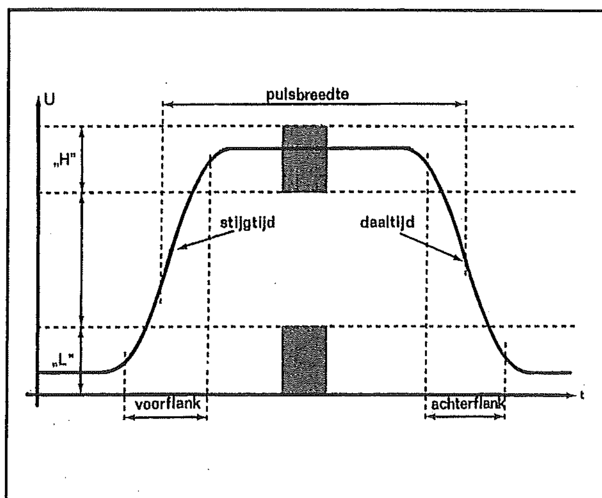
Zonder koeling zou de processor binnen de minuut letterlijk smelten! Die noodzakelijke intensieve koeling stelt de systeembouwers voor nogal wat problemen. Ener-

LEES OOK:

geen verwijzingen



zijds dwingt de moordende concurrentie tot het op de markt brengen van steeds goedkopere systemen, anderzijds dwingt het moderne intensief PC-gebruik tot steeds uitgebreidere koeling. De moderne koelsystemen zijn peperduur en dus wordt nét dat geïnstalleerd dat nét voldoende koelt om de processor in leven te houden.



Figuur 3/19.14-2: De niet oneindig kleine schakeltijden zijn een belangrijke oorzaak van de warmteontwikkeling in een processor.

Schakelen kost ook energie

De zeer hoge kloksnelheden van de moderne processor dragen ook nog op een andere manier bij aan de warmteontwikkeling in de chip. Processoren werken digitaal, alle transistoren schakelen dus hun uitgang naar "L" of naar "H", zie figuur 3/19.14-2. Maar dat schakelen gaat niet in een oneindig korte tijd. Zelfs met de moderne voor- en achterflanken die onder de ns liggen, zal de uitgangspuls van een transistor dus een bepaalde tijd tussen de "L" en de "H" "zweven".
Een transistor in een processor kan zich dus in drie toestanden bevinden:

- De transistor staat open
  Er vloeit dan geen stroom door het onderdeel hetgeen er per wiskundige definitie op neer komt dat er ook géén vermogen in het onderdeel wordt verbruikt, de warmtedissipatie is dus nul.
- De transistor staat gesloten
  Er vloeit nu wél stroom door het onderdeel en het vermogen dat in de transistor wordt opgewekt hangt af van de spanning over de transistor en van de inwendige geleidingsweerstand van het onderdeel. Beide worden tegenwoordig zo laag mogelijk gemaakt. Moderne processoren werken al men een zogenoemde "core"-spanning van minder dan 2,5 V. Er wordt in deze toestand dus wel wat warmte opgewekt, maar dat is minimaal.
- De transistor zit in de voor- of de achterflank van het schakelen
  Op dát moment vloeit er een vrij grote stroom door het onderdeel, terwijl de inwendige weerstand bezig is om te schakelen van (theoretisch) oneindig hoog naar oneindig laag (of vice versa). Er wordt nu dus een behoorlijke hoeveelheid warmte gegenereerd.

Dit alles is géén groot probleem als de schakelsnelheid van de transistoren vrij laag is. Als de "L"- en "H"-perioden van het uitgangssignaal bijvoorbeeld gemiddeld 1 ms lang zijn en de schakeltijden bedragen 10 ns, dan zal de transistor maar één tienduizendse van de tijd in een situatie verkeren, waarin veel warmte wordt gegenereerd. Gedurende de lange "L"- en "H"-perioden kan het onderdeel de gegenereerde warmte gemakkelijk afstaan aan de chip, de behuizing van de processor en de buitenwereld. Als echter de kloksnelheid wordt opgeschroefd tot in het GHz-gebied, dan zullen de voor- en achterflanken een zeer groot deel van de totale pulstijd in beslag nemen. Met andere

woorden: de transistor zit bijna steeds in een situatie waarin veel warmte wordt gegenereerd. Zonder speciale koeltechnieken zou de warmte niet kunnen worden afgevoerd en zou de processor sneuvelen.

Aan- en uitschakelen is op termijn dodelijk
Het gevolg is dat de processor opwarmt tot de grens van wat het onderdeel nog nét kan verdragen. Op zich géén probleem, ware het niet dat een andere natuurwet parten speelt. Iets dat opwarmt zet uit, iets dat afkoelt krimpt. Als een PC dus vaak aan en uit wordt gezet zal de peperdure processor voortdurend afkoelen tot kamertemperatuur en nadien weer opwarmen tot een temperatuur van vér boven de honderd graden. Door het steeds opnieuw uitzetten en krimpen van de behuizing ontstaan mechanische spanningen in het onderdeel die op termijn tot beschadiging leiden. Beschadigingen die meestal het levenseinde van dit duurste onderdeel van een PC inluiden.

Water koelt beter dan lucht
Het zou dus heel verstandig zijn een processor niet iedere keer opnieuw tot zijn maximale bedrjfstemperatuur op te warmen. Lagere bedrijfstemperaturen hebben minder mechanische spanningen tot gevolg. Maar dan zouden de radiateuren nog groter en nog duurder moeten worden dan ze nu al zijn.
Er is echter een voor de hand liggende oplossing. De radiateur in een auto koelt het motorblok in de meeste gevallen niet met lucht maar met water. Het gloeiend hete water wordt rondgepompt en in de radiateur met geforceerde luchtstroming afgekoeld. Waarom dit systeem niet toepassen op een processor? Water heeft immers een veel grotere warmtecapaciteit dan lucht, met als gevolg dat water veel beter in staat is de warmte-ontwikkeling in de processor af te voeren.

Een lekker koel en stil systeem
Waterkoeling heeft nog andere voordelen. Aangezien de warmte bij een goed aangelegd systeem wordt afgevoerd vanuit de kast naar buiten zal de temperatuur in de kast ook dalen, de processor is immers de belangrijkste hittebron in de kast. Waterkoeling leidt dus tot een lagere temperatuur van de gehele computer. Een factor waar vaak minder op wordt gelet, is de geluidsproductie. Doordat het aantal fans om de PC te koelen flink kan worden teruggebracht (tot 1 of 2) zal het geluidsniveau dramatisch afnemen. Een beter gekoeld systeem met minder geluidsproductie dus!

Een verlate één april grap?
Écht niet! Waterkoeling van processoren is een serieuze zaak aan het worden en verschillende fabrikanten brengen spullen op de markt waarmee men, als men tenminste geen twee linkerhanden heeft, in een mum van tijd de PC van waterkoeling voorziet. Tegenwoordig zijn er zelfs handige kits op de markt, waar alle noodzakelijke onderdelen in zitten én een uitgebreide handleiding. Waterkoeling installeren is tegenwoordig bijna nét zo gemakkelijk als een of andere nieuwe kaart monteren.

Het systeem

Wat is er minimaal nodig voor het inbouwen van een watergekoeld systeem? Dat valt wel mee:
- Een koelblok;

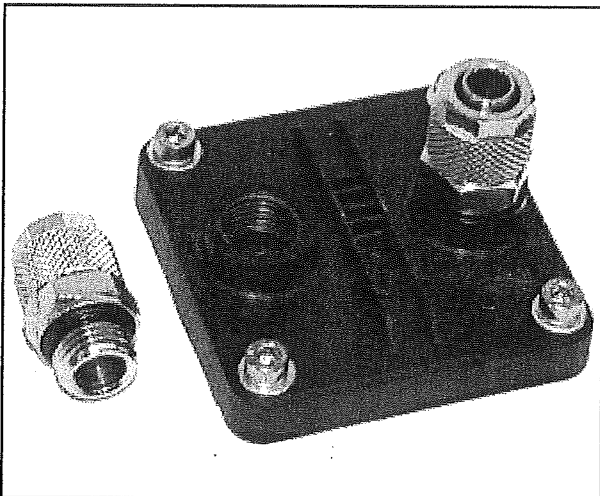
- Een waterslang;
- Een waterpomp;
- Een waterreservoir;
- Een radiateur.

In de volgende paragrafen worden deze onderdelen even nader bekeken.

Het koelblok

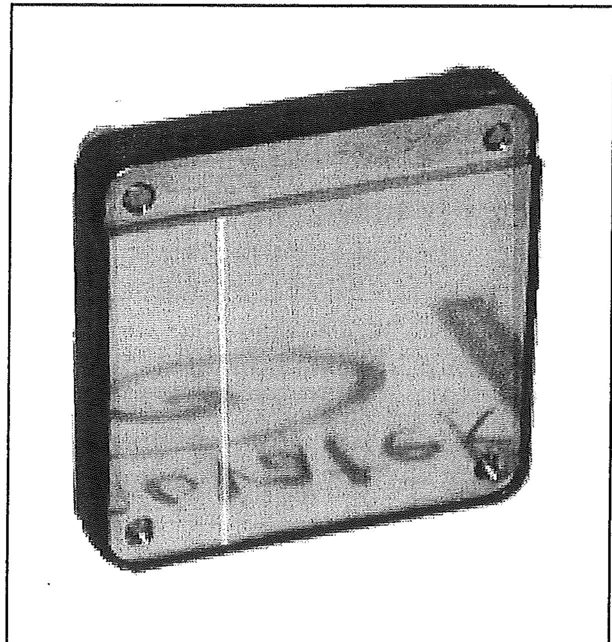
De warmte van de processor afvoeren

Het koelblok komt in de plaats van de bestaande koelplaat op de processor te zitten en heeft als doel de warmte van de processor af te geven aan het water dat door het koelblok stroomt. Een watergekoeld blok is veel kleiner dan de grote koelplaat mét ventilator die nu op de processor zit.



Figuur 3/19.14-3: Een voorbeeld van een koelblok met waterkoeling.

Koelblokken zijn van koper of aluminium. Koper geleidt de warmte iets beter dan aluminium, maar een goed ontworpen aluminiumblok neemt meer hitte van de processor weg dan een slecht koperblok. Maar, hoe dan ook, elk waterblok functioneert véél beter dan welke luchtkoeler dan ook! Koelblokken zijn in diverse uitvoeringen leverbaar en natuurlijk ook voor alle processormodellen. In figuur 3/19.14-3 wordt als voorbeeld de "InnovaFlat Flow-O-Matic" van Innovatek voorgesteld. Dit koelblok bestaat uit een koperblok van 60 bij 60 mm^2 en is slechts 18 mm hoog. Aan de bovenzijde bevinden zich twee klemmen waarmee men de waterslangen voor de aan- en afvoer van het koelwater absoluut lekvrij op het onderdeel kan bevestigen. De onderzijde van het koelblok is uiteraard absoluut glad afgewerkt, zie figuur 3/19.14-4, want deze bodemplaat wordt uiteraard op de behuizing van de processor bevestigd.



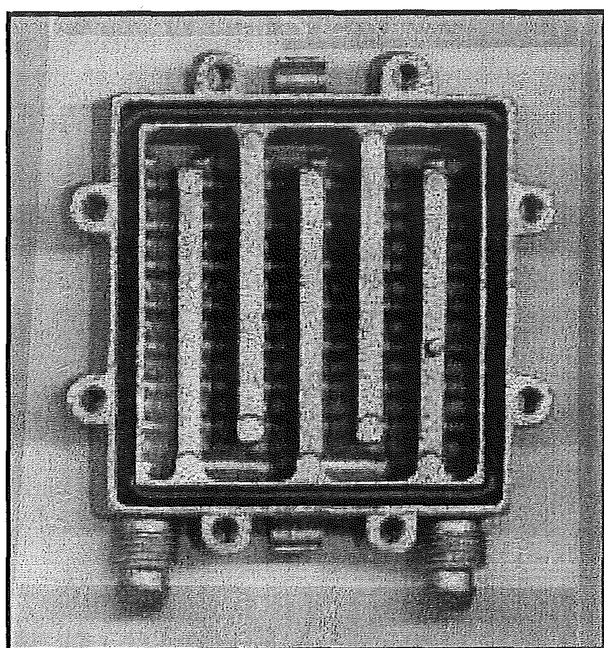
Figuur 3/19.14-4: De volledig glad afgewerkte onderzijde van het koelblok wordt op de behuizing van de processor bevestigd.

Op een lange waterweg komt het aan

Als men zo'n koelblok zou openslopen zou men vaststellen dat het koperen of aluminium blok in een zigzagvormig patroon is uitgefreesd, zie figuur 3/19.14-5.

19.14 Waterkoelingstechnologie voor snelle processoren

Dat is logisch, want het water moet zo lang mogelijk door het blok stromen om zoveel mogelijk warmte af te voeren.



Figuur 3/19.14-5: Een opengesloopt koelblok: het koelwater stroomt zigzaggend door het metaal zodat zoveel mogelijk warmte wordt geabsorbeerd.

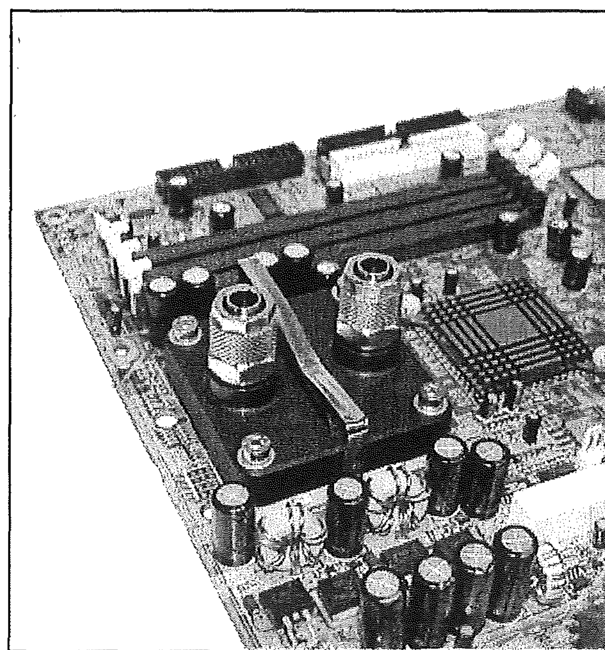
Deze kanalen hebben wél als nadeel dat de wrijving tussen het water en het "kanaal" vrij hoog wordt en er dus flink gepompt moet worden om de waterstroom op gang te houden.

Hier zit het kwaliteitsverschil

Er bestaan aanzienlijke prijsverschillen tussen de aangeboden koelblokken. De zorg die is besteed aan de interne afwerking van het koelblok heeft hier veel mee te maken. Soms zijn de kanalen tamelijk slordig uitgefreesd met niet al te gladde wanden (zoals in het voorbeeld van figuur 3/19.14-5), soms zijn deze koelgangen werkelijk perfect glad. Het zal duidelijk zijn dat in het eerste geval de doorstroming van het koelwater nogal wordt gehinderd. Tegenwoordig zijn er zelfs fabrikanten die gebruik maken van volledig gegoten aluminium blokken, waarbij de afwerking van het inwendige kanaal werkelijk perfect is. Deze aluminium koelers zijn echter peperduur.

Op de processor monteren

Het monteren van een watergekoeld koelblok op de processor gaat net zo gemakkelijk als het monteren van een klassieke radiateur. Men moet er natuurlijk wél voor zorgen dat het juiste type voor de processormodel in huis wordt gehaald. In figuur 3/19.14-6 is te zien hoe zo'n koelblok met de gebruikelijke klem aan de voet van de processor wordt gevestigd.

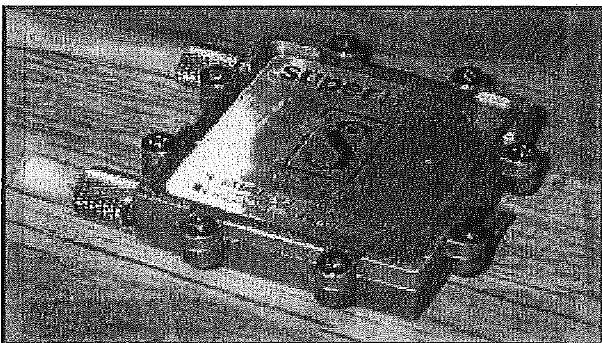


Figuur 3/19.14-6: Het monteren van een watergekoeld koelblok op de processor.

Diverse uitvoeringen

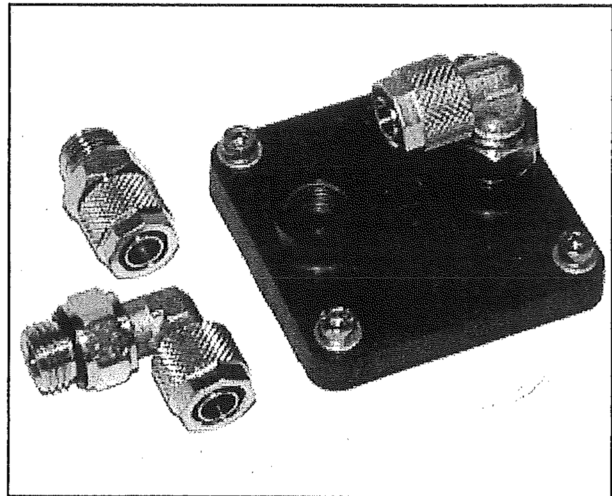
Ieder koelblok is, dat kan niet anders bij een radiateur, voorzien van twee aansluitnippels voor de aan- en afvoer van het

koelwater. Er zijn koelblokken waar deze nippels aan de bovenzijde zitten en koelblokken waar deze aan de zijkant zitten. Het tweede model heeft als voordeel dat het water gemakkelijker door de zigzagvormige kanalen in het blok kan stromen. Het groter voordeel van het eerste model is natuurlijk dat de aan- en afvoerslangetjes gemakkelijker op het koelblok worden aangesloten. In figuur 3/19.14-7 is als voorbeeld het koelblok H2O van Senfu voorgesteld, waar de slangetjes aan de zijkant worden aangesloten.



Figuur 3/19.14-7: Een alternatieve uitvoering van een koelblok met de aansluitnippels aan de zijkant van het koelblok.

Voor de meeste modellen met aansluitnippels aan de bovenzijde van het koelblok zijn adapters te koop, zodat men de slangetjes ook evenwijdig aan de printplaat van het moederbord kan aansluiten. In figuur 3/19.14-8 is dit voorgesteld voor het reeds eerder aangehaalde koelblok van de firma Innovatek. Dit kan handig zijn als men besluit niet alleen de processor te voorzien van waterkoeling, maar ook de grafische processor op de grafische kaart. Deze zit in een uitbreidingsslot en er is daar uiteraard geen ruimte op de slangetjes loodrecht op het koelblok aan te sluiten. In figuur 3/19.14-9 is een speciaal voor de waterkoeling van de grafische processor ontwikkeld koelblok voorgesteld van de firma 2Cool. Dit koelblok is extreem dun en de aansluitnippels nemen erg weinig ruimte in. Een en ander heeft tot gevolg dat er maar één slot verloren gaat.



Figuur 3/19.14-8: Adapters voor het radiaal aansluiten van de koelslangen op een axiaal koelblok.

Corrosie!

Het koelblok is spuitgegoten uit aluminium of gefreesd uit koper. Dit zorgt voor een hoge warmte-overdracht, hetgeen een wenselijke eigenschap is in een koelinstallatie. Het is echter onvermijdelijk dat bij gebruik van deze metalen galvanische corrosie optreedt. Dit beperkt de levensduur van het koelblok. Sommige waterkoelers gebruiken antivries om dit effect tegen te gaan. Sommige fabrikanten adviseren kant-en-klare koelvloeistof uit de autobranche, mits geschikt voor aluminium.

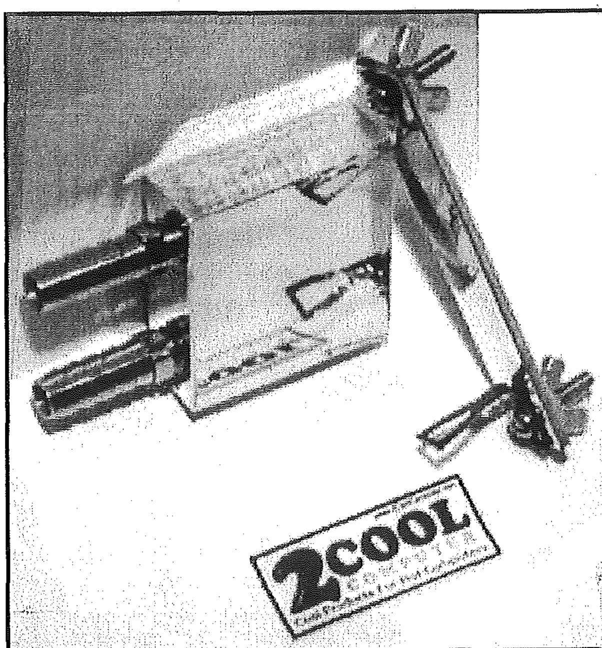
Fabrikanten

- 2Cool

  2Cool maakt een behoorlijk groot aluminium waterblok. In het blok zijn twee kanalen geboord, met een doorlaat

van 6 mm. Het blok is geheel hoogglans gepolijst.

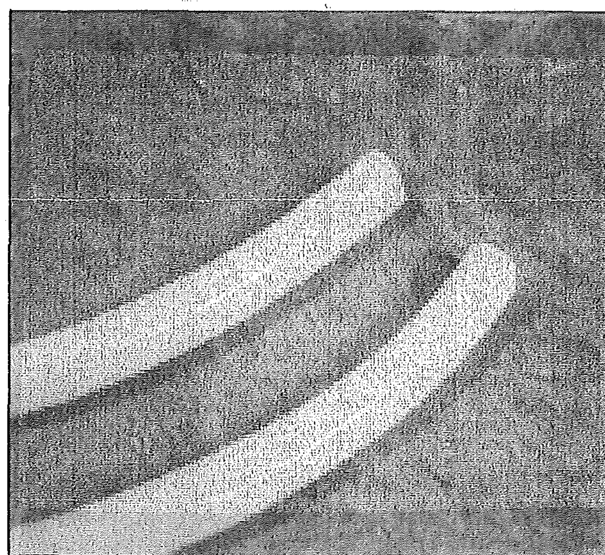
- Aquacool
  Aquacool maakt koelblokken, welke geheel voorzien zijn van een verguld laagje. Dit gaat corrosie tegen. De aansluitingen zijn 6 mm. Dit blok komt met een keur aan klemmen, om de koeler op vrijwel alle gangbare processoren te kunnen gebruiken. Deze koeler wordt enkel in een set samen met slang en pomp verkocht. De Deluxe versie komt in geschenkverpakking, met als extra een videokaart koeler.
- Danger-Den
  Danger-Den maakt koperen koelblokken, welke met een frees voorzien worden van een kanalenstelsel. Hierna wordt een deksel op het koelblok gesoldeerd. Op de Danger-Den site is te zien hoeveel druk het blok aankan, voordat het kapot gaat. Dit is behoorlijk indrukwekkend.



Figuur 3/19.14-9: Een extreem dun koelblok voor montage op printen die in de uitbreidingssloten zitten.

De waterslang

Aan- en afvoeren van het koelwater
Om het koelwater in het koelblok te krijgen en weer af te voeren, kan men slang of buis gebruiken. Slang is uiteraard de gemakkelijkste optie, u bent immers waarschijnlijk geen gediplomeerd loodgieter. Slang is in diverse maten, materialen en kwaliteit te koop. De maat moet uiteraard zijn afgestemd op de nippels van het koelblok, de pomp en de radiateur. Koop kwalitatief hoogwaardige slang, het zal wel niet noodzakelijk zijn om te waarschuwen tegen de gevaren van lekkages in een PC! Deze slangen lijken een beetje op normale aquariumslang, maar volgens de fabrikant is de kwaliteit onvergelijkbaar veel beter, zie figuur 3/19.14-10.



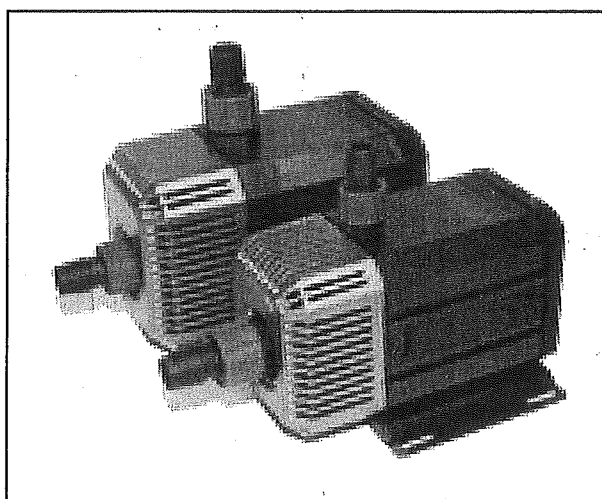
Figuur 3/19.14-10: De waterslangetjes zorgen voor het transport van het koelwater tussen de verschillende onderdelen van het systeem.

19.14 Waterkoelingstechnologie voor snelle processoren

Hoe flexibeler, hoe beter

Een flexibele slang, zoals siliconenslang, is makkelijker aan te leggen, en geeft minder kans op lekkage, omdat een flexibele slang minder snel van een koppeling afschiet. Om te voorkomen dat een slang van de koppeling afschiet, kan de slang erop vastgeklemd worden met een tiewrapje. In het meest gunstige geval is het koelblok voorzien van nippels in paddestoelvorm, waar de slang stevig overheen gedrukt wordt. Zorg er in ieder geval voor dat de slang klein genoeg is om uit zichzelf stevig om een aansluiting heen te klemmen. Eventueel kan door het eind van de slang warm te maken deze makkelijker over een aansluiting geschoven worden. Let er ook op dat de slang zich in de bochten niet platknijpt. Dit is vooral van toepassing op goedkopere, stuggere slang. Soepele siliconenslang heeft dit nadeel niet, maar is wel duurder.

In figuur 3/19.14-11 zijn twee pompen voorgesteld van het fabrikaat Eheim. Dat zijn géén aquariumpompen maar échte centrifugaalpompen met een debiet van respectievelijk 5 en 10 liter per minuut. Bovendien zijn deze pompen extreem geluidsarm, iets dat van aquariumpompen niet per definitie gezegd kan worden.



Figuur 3/19.14-11: Deze pompmodellen van AHEIM werken volgens het centrifugaal principe.

De pomp

Op pompen komt het aan

De pomp is het belangrijkste onderdeel van een waterkoeling, want als dit onderdeel het begeeft is de ramp niet te overzien. Er bestaan uiteraard diverse modellen en uitvoeringen. De pomp kan een eenvoudig model zijn dat op 12 V draait, of een model dat op de netspanning van 230 V loopt. Sommige leveranciers van waterkoelingsystemen doen er niet moeilijk over. Pompen die voor aquaria worden gebruikt zijn blijkbaar uitstekend in staat dit alternatief klusje te klaren. Let er echter wél op dat nogal wat materiaal afkomstig is uit Amerika of Taiwan en dat netgevoede pompen daar op 110 V draaien.

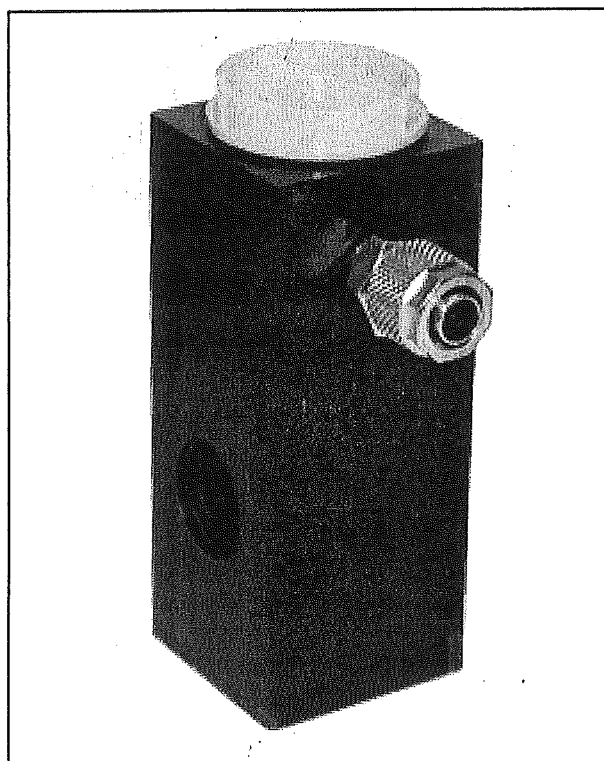
Impeller-pompen (de meeste aquariumpompen) kunnen probleemloos afgeknepen worden met een instelbare knijper op de slang als de stroomsnelheid te hoog is. Meestal is dit echter niet noodzakelijk, want bij een waterkoeling is de pomp meestal al behoorlijk afgeknepen door kleine slangen, kleine doorvoeren en een hoge stromingsweerstand. Dit is niet schadelijk voor de pomp.

Belangrijke opmerking

Laat, voor de pomp aan te zetten, eerst het pomphuis vollopen met water. Dit voorkomt dat de pomp droog loopt en kapot gaat. Dit is doodsoorzaak nummer één voor waterpompen!

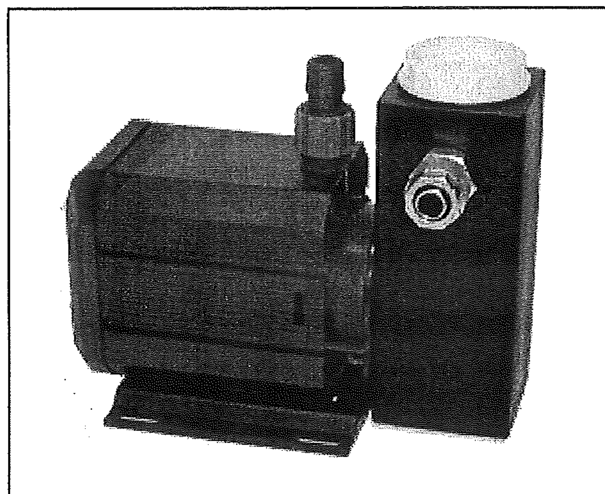
Het reservoir

Een voorraadje water is wél zo handig
Natuurlijk moet het systeem worden gevuld met water. Daarvoor zijn handige reservoirs op de markt, de op de achterzijde van de PC-kast worden gemonteerd. In figuur 3/19.14-12 wordt als voorbeeld de "Ausgleichsbehalter" van AHEIM gepresenteerd. Dit tankje is gemaakt uit geloxeerd aluminium en is aan de bovenzijde voorzien van een groot afgedicht gat voor het bijvullen van het systeem.



Figuur 3/19.14-12: Een voorraadvat zorgt voor voldoende watervoorraad in het koelsysteem.

Bij sommige systemen worden pomp en reservoir tot één geheel samengebouwd, zie figuur 3/19.14-13. Kwestie van voorkeur, maar of dat érg handig is, is maar de vraag!



Figuur 3/19.14-13: Bij het systeem van Eheim worden pomp en voorraadvat tot één geheel samengebouwd.

De radiateur

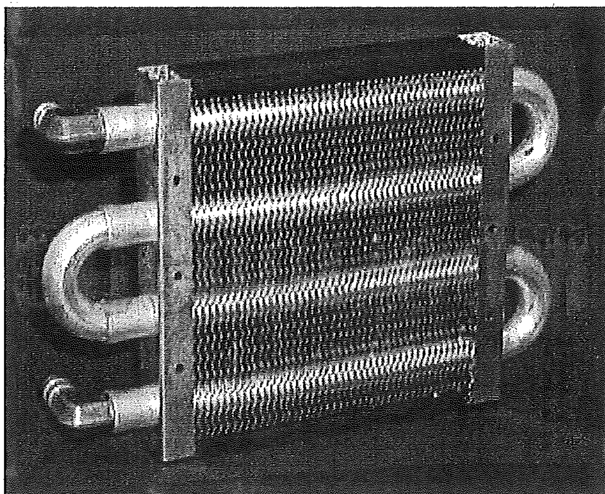
Heet water wordt gekoeld
De radiateur moet het warme water weer koelen nadat het door de processor is verwarmd. Een radiateur wordt in de slang opgenomen tussen het koelblok en het reservoir, of tussen het reservoir en het koelblok.

Diverse uitvoeringen
Radiateuren zijn er in alle maten, soorten en uitvoeringen. In de eenvoudigste uitvoering wordt het water door een zigzagvormige metalen pijp gevoerd, die voorzien is van koelribben. Het geheel kan tegen de achterzijde van de systeemkast worden bevestigd en het is de bedoeling dat de circulerende lucht voor voldoende koeling zorgt, zie figuur 3/19.14-14.

Radiateur met geforceerde koeling
Toch is het niet verstandig een dergelijke radiateur aan te schaffen. Als men tóch

19.14 Waterkoelingstechnologie voor snelle processoren

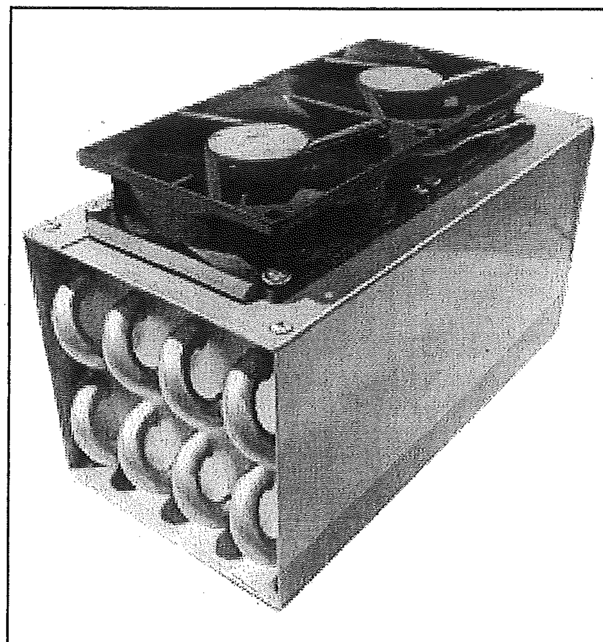
bezig is kan men meteen het beste van het beste installeren. Kies dus een radiateur die is voorzien van een of twee ventilatoren, die zorgen voor een geforceerde warmte-overdracht tussen het koelwater en de lucht. In figuur 3/19.14-15 wordt de "Dual Fan Radiator" van de firma Senfu voorgesteld. Bij dit model wordt het hete water door een uitgebreid pijpensysteem gevoerd dat via koelribben en twee ventilatoren op kamertemperauur wordt gehouden. Zoals uit de foto van figuur 3/19.14-19 blijkt, hetzelfde model maar dan van de andere kant bekeken, is deze radiateur uitgebreid voorzien van koelribben waar de koellucht doorheen stroom. Een en ander staat garant voor een optimale warmte-uitwisseling tussen het koelwater en de omgevingslucht.



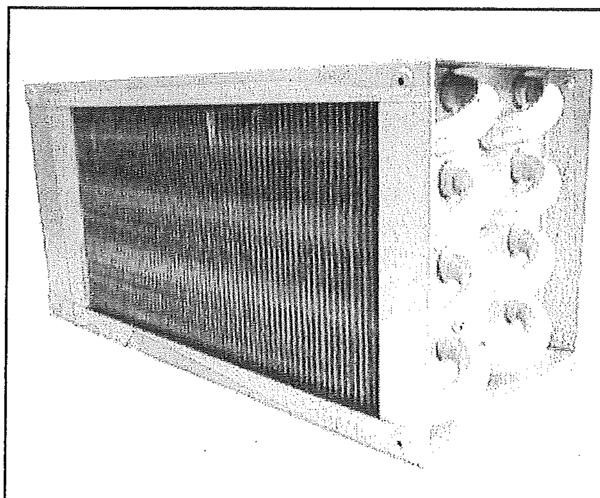
Figuur 3/19.14-14: De eenvoudigste uitvoering van een radiateur: de warmtewisseling tussen water en lucht gebeurt door natuurlijke convectie.

Plaatsing van de radiateur

De radiateur kan op drie manieren geplaatst worden: in de ingaande luchtstroom de kast in, in de uitgaande luchtstroom de kast uit, of geheel buiten de kast. Het voordeel van de laatste twee oplossingen is dat de warmte van de processor niet in de kast terecht komt, waardoor verdere fans waarschijnlijk overbodig zijn.



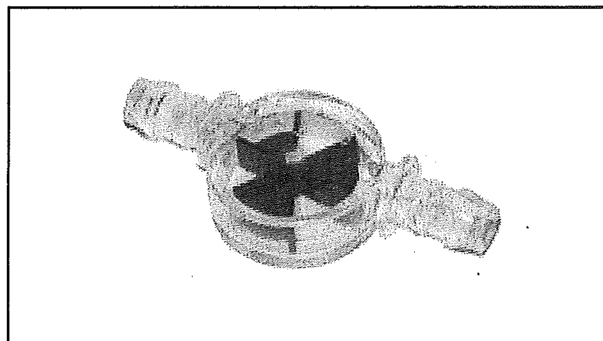
Figuur 3/19.14-15: De "Dual Fan Radiator" van Senfu.



Figuur 3/19.14-16: De koelribben in de "Dual Fan Radiator" van Senfu zorgen voor een maximale warmte-uitwisseling tussen water en lucht.

19.14 Waterkoelingstechnologie voor snelle processoren

Het mooiste is dus de radiateur te plaatsen in de luchtstroom naar buiten. Door toepassing van waterkoeling zal de temperatuur in de kast niet hoog zijn. Warmte van de processor wordt niet de kast alsnog ingeblazen, maar wordt de kast uitgeblazen.
Door alle onderdelen van het systeem in de kast te houden is het systeem natuurlijk wél minder gevoelig voor storing.



Figuur 3/19.14-17: De flowmeter van het systeem: een duur woord voor een eenvoudig en goedkoop onderdeeltje dat in feite onmisbaar is.

Extra's

Dat was het, of toch niet?
In principe bestaan een waterkoelingssysteem dus niet uit erg veel onderdelen. Toch is er één extra dat in de meeste bouwpakketten ontbreekt en onmisbaar is.

Een flowmeter
Doet hij het wél of doet hij het niet? Een belangrijke vraag, want als het mooi waterkoelingsysteem in de fout gaat kost dat de dure processor. Vandaar dat ten stelligste wordt aangeraden een flowmeter in het systeem op te nemen. Het principe is bekend van de benzinepomp. Een vaantje zit in de stromende vloeistof. Draait het vaantje, dan pompt het systeem water naar de hete processor. Staat het vaantje stil, dan is er iets mis en moet de PC onmiddellijk worden uitgeschakeld.
In figuur 3/19.14-17 wordt de "Flow-O-Matic" van Kavan voorgesteld. Het apparaatje is voorzien van aansluitingen met een diameter van 9,8 mm en wordt simpelweg ergens in een van de slangen opgenomen. Het is uiteraard wél de bedoeling dat dit apparaatje ergens goed zichtbaar wordt gemonteerd, want anders heeft men er nog niets aan.

Thermometers
Door verschillende leveranciers van waterkoelingssystemen worden digitale thermometers aangeboden, die de temperatuur van het koelwater bewaken. Een veel beter, maar ook veel duurder systeem dan de simpele flowmeter. Er zijn uiteraard exemplaren die met de PC communiceren via een seriële poort of een USB-connector. De bijgeleverde software zorgt dat er een alarmvenster op de monitor verschijnt op het moment dat de temperatuur van het koelwater boven een bepaalde grens stijgt. Een héél luxueus systeem dat de dure ingewanden van de moderne PC optimaal beschermt!

Fabrikanten en leveranciers

Fabrikanten
Er zijn inmiddels heel wat fabrikanten in de waterkoeling gesprongen. Vreemd genoeg hebben wij geen enkele Europese fabrikant kunnen ontdekken. De grote

spelers op deze markt zijn Amerikaanse en Taiwaneese bedrijven, zoals:
- Senfu, www.senfu.com.tw;
- Aquacool, www.aquacool.com;
- Danger-Den, www.dangerden.com;
- 3D-Cool, www.3dcool.com.

Nederlandse leveranciers
Diverse bedrijven leveren onderdelen en complete set's voor het inbouwen van een waterkoelingsysteem.
- De Regenboog
  Deze grote PC-leverancier uit het zuiden des lands stelt op de site www.regenboog.nl een uitgebreid assortiment onderdelen en ook en paar complete set's voor.
- OC Shop
  Dit bedrijf is gespecialiseerd is allerlei buitenissigheden in en rond de PC. Waterkoeling mag dus niet ontbreken. Op www.oc-shop.nl kan men on-line waterkoelingen bestellen.
- Conrad Elektronica
  Dit aan iedere elektronicus bekende filiaal van de grootste Duitse elektronica leverancier heeft géén losse onderdelen in de verkoop, maar maakt het gemakkelijk door een complete set aan te bieden die op alle processoren past. Deze set is samengesteld door het Duitse bedrijf Innovatek uit onderdelen van diverse merken. Voor EURO 259,00 kan men op www.conrad.nl deze koelset on-line bestellen.

3/19.15
De Agent-technologie van Microsoft

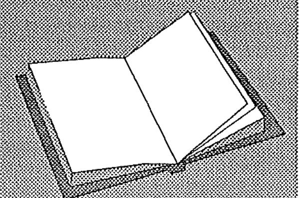
Inleiding

Een uitbreiding van de Office Assistent
Wie werkt met een zeer recente versie van het Office-pakket van Microsoft weet in feite al wat de Agent-technologie voorstelt. In de Office-onderdelen kan men namelijk in het helpmenu de optie "De Office-assistent weergeven" aanvinken, waarna een kleine, bewegende paperclip op het scherm verschijnt die via tekstballonnen de weg wijst door de vele opties van de Office-programma's. Deze assistent, zie figuur 3/19.15-1, is in feite niets meer of minder dan een miniatuur-Agent, zij het dat hij niet kan praten.
Microsoft heeft deze technologie uit het Office-pakket losgemaakt en er een op zichzelf staande technologie van gemaakt: de Agent-technologie. De eerste stappen op dit pad werden reeds in 1998 gezet, maar het is eigenlijk sinds een halfjaar dat deze technologie volwassen aan het worden is en de mogelijkheden zijn uitgebreid.
Agents kunnen niet alleen bewegen over het scherm, maar kunnen via de geluidskaart ook teksten uitspreken en reageren op teksten die via een microfoon worden ingesproken. Het is zelfs mogelijk een Agent volledig lipsynchroon een tekst te laten uitspreken!

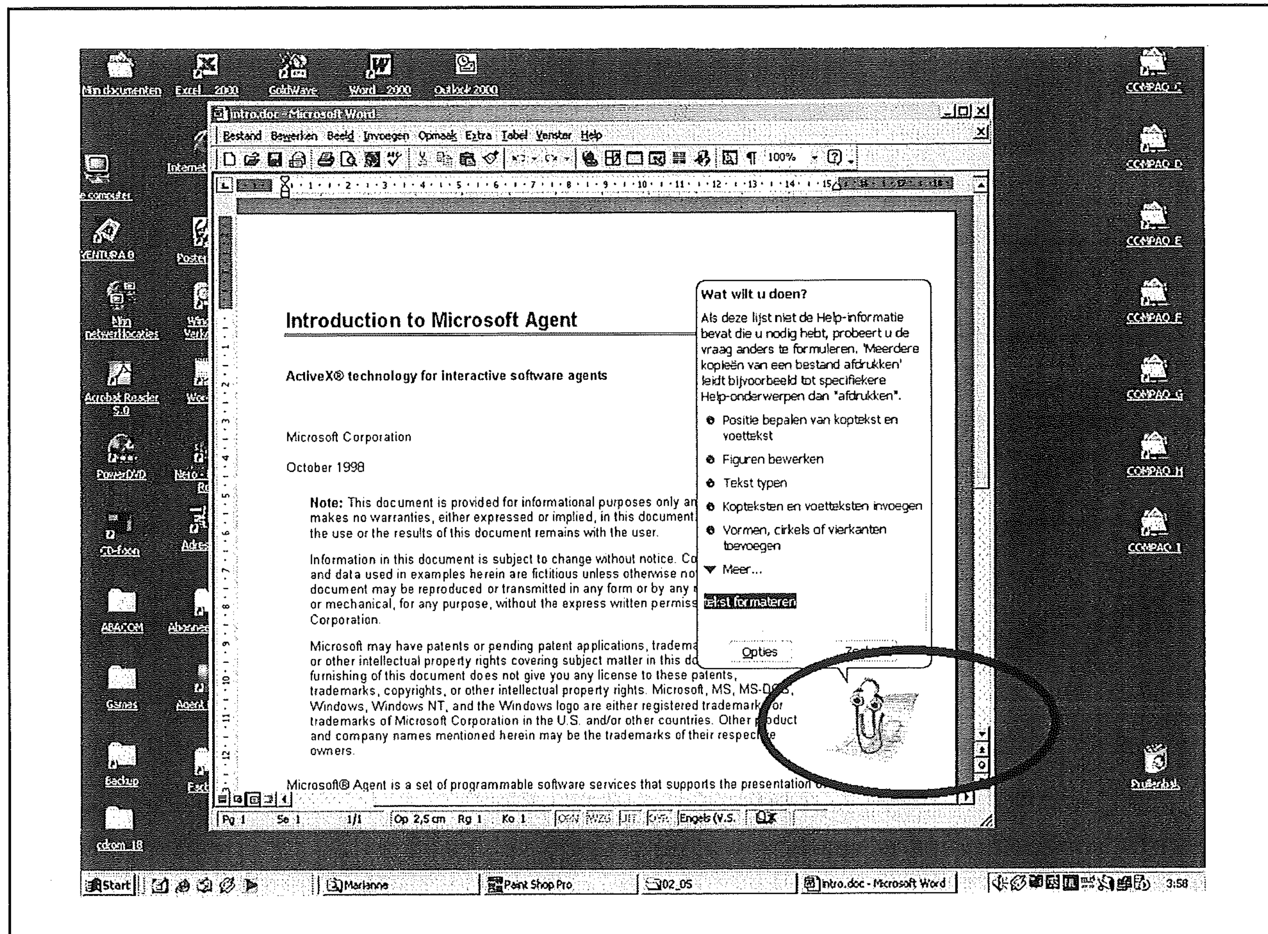
Een in het bedrijfssysteem geïntegreerde technologie
De Agent-technologie is geen Windows-applicatie, maar wordt een onderdeel van het bedrijfssysteem. In sommige versies van Windows 2000 en XP zijn reeds bepaalde componenten van deze technologie geïntegreerd. Maar men kan er zelfs in Windows 95 gebruik van maken. Als men de noodzakelijke componenten in het systeem heeft geïnstalleerd, merkt men er nog niets van. De Agents komen maar eerst tot leven als de een of andere applicatie er gebruik van maakt.

De Agents vormen de basis
Basis van de Agent-technologie vormen de Agents. Geanimeerde figuren of karakters die ergens op het scherm tevoorschijn komen als zij door een applicatie worden aangeroepen. De afmetingen van deze Agents zijn vrij groot.

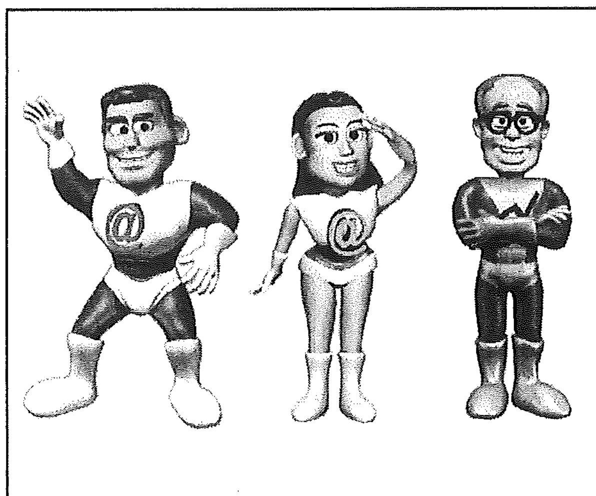
LEES OOK:
Hoofdstuk 5/14.2.1



19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-1: De Agent-technologie is in beperkte vorm aanwezig in de recente versies van Office.



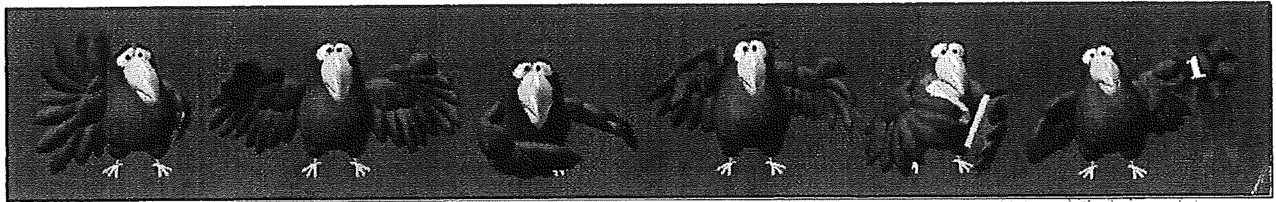
Figuur 3/19.15-2: Drie mooie Agents leveren het bewijs van de gedetailleerdheid van de karakters.

Ze worden weergegeven in een onzichtbaar venster met als maximale afmetingen 128 bij 128 pixels.
Een en ander heeft tot gevolg dat deze Agents heel mooi en gedetailleerd kunnen worden getekend. Kijk maar eens naar figuur 3/19.15-2, waarin drie mooie Agents als voorbeeld worden gepresenteerd.

De samenstelling van een Agent

Een Agent bestaat uit animaties. Een animatie is een basisbeweging die de Agent kan uitvoeren, zoals “Luisteren”, “Groeten”, “Naar links kijken”, “Naar boven kijken”, “Verbaasd kijken” enzovoort.

19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-3: Zes animaties van de Microsoft Agent Peedy.

Deze animaties zijn hier met hoofdletters geschreven omdat ze ook met deze namen worden aangeroepen door applicaties, maar dan uiteraard in het Engels. Goede Agents kunnen meer dan vijftig verschillende animaties bevatten. Elke animatie is weer samengesteld uit vaak honderden frames. Dat zijn statische BMP-illustraties, die, net zoals in een film of in een animated GIF, de bewegingen van de Agent tot stand brengen doordat ze snel achter elkaar in het venster van de Agent worden weergegeven. Het zal dan ook geen verbazing wekken dat een Agent-bestand vaak MB's groot is. In figuur 3/19.15-3 zijn als voorbeeld zes animaties van een van de door Microsoft ontwikkelde Agents voorgesteld.

Een universele technologie

Inleiding

De Agent-technologie is een zeer universele technologie. Men kan Agents bijvoorbeeld aanroepen vanuit HTML-pagina's door gebruik te maken van JavaScripts. Men kan Agents toepassen in Visual Basic, Visual C++, Visual Basic for Applications (VBA) en JScript. Er zijn inmiddels programmeeromgevingen op de markt waarmee men Agents kan inbedden in PowerPoint-presentaties. Vanwege de samenwerking tussen Microsoft en het Belgische bedrijf Lernout & Hauspie, dat de spraakengines heeft ontworpen, kan men Agents goed verstaanbaar Nederlands laten praten. In het Frans of Duits? Geen probleem, want ook voor deze talen, en overigens voor een heleboel andere, staan spraakengines ter beschikking. Kortom, er is maar weinig dat met de Agenttechnologie niet kan.

Toegang tot de Agents

Er bestaan twee manieren waarop men toegang krijgt tot een Agent. Men kan het volledige bestand, dat dus alle animaties met al hun frames bevat, op de harde schijf installeren. Dat gebeurt onder de vorm van ACS-bestanden, die vaak meer dan 2 MB groot zijn. Als men een applicatie installeert die gebruik maakt van de Agent-technologie, dan laadt die applicatie op de juiste momenten de juiste animaties uit dit bestand. Men kan de Agenttechnologie ook gebruiken op Internet pagina's. Maar het zal wel duidelijk zijn dat het niet praktisch is als een bezoeker van een website eerst een volledig Agentkarakter moet downloaden. Vandaar dat Microsoft hiervoor iets tijdbesparends heeft bedacht.

Op de webserver http://agent.microsoft.com/agent2/chars staan alle animaties van de vier door Microsoft ontworpen Agents als losse kleine bestanden. Een ontwerper van Internet pagina's kan met LOAD- en GET-instructies toegang krijgen tot die bestanden en een welbepaalde animatie downloaden.

19.15 De Agent-technologie van Microsoft

Zo'n 75 verschillende snacks of snackvariaties!

onze frikadellen worden in eigen keuken bereid uit de fijnste ingrediënten

Frikadel	1,00
Frikadel speciaal	1,45
Frikadel speciaal zuurvlees	1,80
Frikadel oriëntal	€ 1,80
Kroket	€ 1,00
Roomkroket	€ 1,25
Goulashkroket	€ 1,10
Kip-champignonkroket	€ 1,10
Kaaskroket	€ 1,10
Groentekroket	€ 1,10
Satékroket	€ 1,25
Champignola (nieuw)	€ 1,25
Bamischijf	€ 1,00
Superbami/nasi	€ 1,15
Rijst-groenteschijf	€ 1,50
Groente-kaasschijf	€ 1,50
Bamibal (eigen keuken)	€ 1,70
Bamsa	€ 1,25
Loempia klein	€ 1,60
Loempia	€ 2,00
Loempia taugeh	€ 2,00

Figuur 3/19.15-4: Op deze moderne Internetpagina van een friture prijst een wijsneuzige vliegende papegaai de producten van de eigenaar in duidelijk uitgesproken Nederlands aan.

Deze worden dan opgeslagen in de Internet cache op de PC van de bezoeker. Omdat zo'n bestand vrij klein is, gaat dat vrij snel. Het enige probleem is dat men hiervoor een gebruikerslicentie bij Microsoft moet aanvragen.

Een geheel nieuwe wereld

Rond het verschijnsel Agent-technologie heeft zich in snel tempo een geheel nieuwe wereld ontwikkeld. Men kan letterlijk honderden applicaties downloaden of kopen die op de een of andere manier gebruik maken van Agents. Men kan door gespecialiseerde bedrijven een eigen bedrijfs-Agent laten ontwerpen. Ook zijn er al tamelijk veel Amerikaanse Internetsites die Agents in hun pagina's hebben geïntegreerd.

Een praktijkvoorbeeld

Stel dat er een friture is die een Internetsite heeft waarop men een menu kan samenstellen, dat vervolgens per e-mail

wordt besteld. "Vroeger" zou, als men op het woord "frikandel" klikte, waarschijnlijk een extra venster openen waarin de friture-eigenaar zijn frikandellen extra aanprees. Als deze man echter gebruik maakt van de Agent-technologie, vliegt er nu opeens een papegaai over het scherm, zie figuur 3/19.15-4, die op het browservenster gaat zitten en in goed verstaanbaar en duidelijk Nederlands zegt: "Onze frikandellen worden in eigen keuken bereid uit de fijnste ingrediënten." Kan een Internetpagina nog dynamischer worden?

Installatie van de Agents-componenten

Inleiding

Zoals reeds geschreven in de inleiding wordt de Agent-technologie een onderdeel van het Windows-systeem. De diverse componenten worden in de Windows-directory geïnstalleerd en zijn, althans volgens Microsoft, daaruit niet meer te verwijderen.

Als men dus om de een of andere reden de Agent-technologie weer kwijt wil, zit er niets anders op dan de harde schijf gedeeltelijk leeg te maken en Windows opnieuw te installeren. Men kan natuurlijk ook, voordat de Agent-technologie wordt geladen, een back-up maken van de Windows-directory.

Het installeren van de Agent-componenten

Microsoft zélf is vrij onduidelijk over de integratie van de Agent-technologie in de verschillende versies van Windows. In Windows 2000 en XP zijn sommige componenten geïntegreerd maar andere niet. Bovendien zijn sommige componenten reeds verouderd. Men kan dus maar beter de volledige installatie uitvoeren, zodat het systeem wordt voorzien van de modernste Agents-componenten.

Wat men in elk geval moet installeren zijn:
- de Microsoft Agent core components;
- minstens één Microsoft Agent Character-bestand;
- de Lernout & Hauspie Truvoice text-to-speech engine, Amerikaans-Engels;
- de SAPI 4.0 runtime support;
- het Microsoft Speech Control Panel.

Met deze vijf componenten kan men alle applicaties die gebruik maken van de Agent-technologie testen en spraakuitvoer in het Amerikaans beluisteren.

Het downloaden van de bestanden

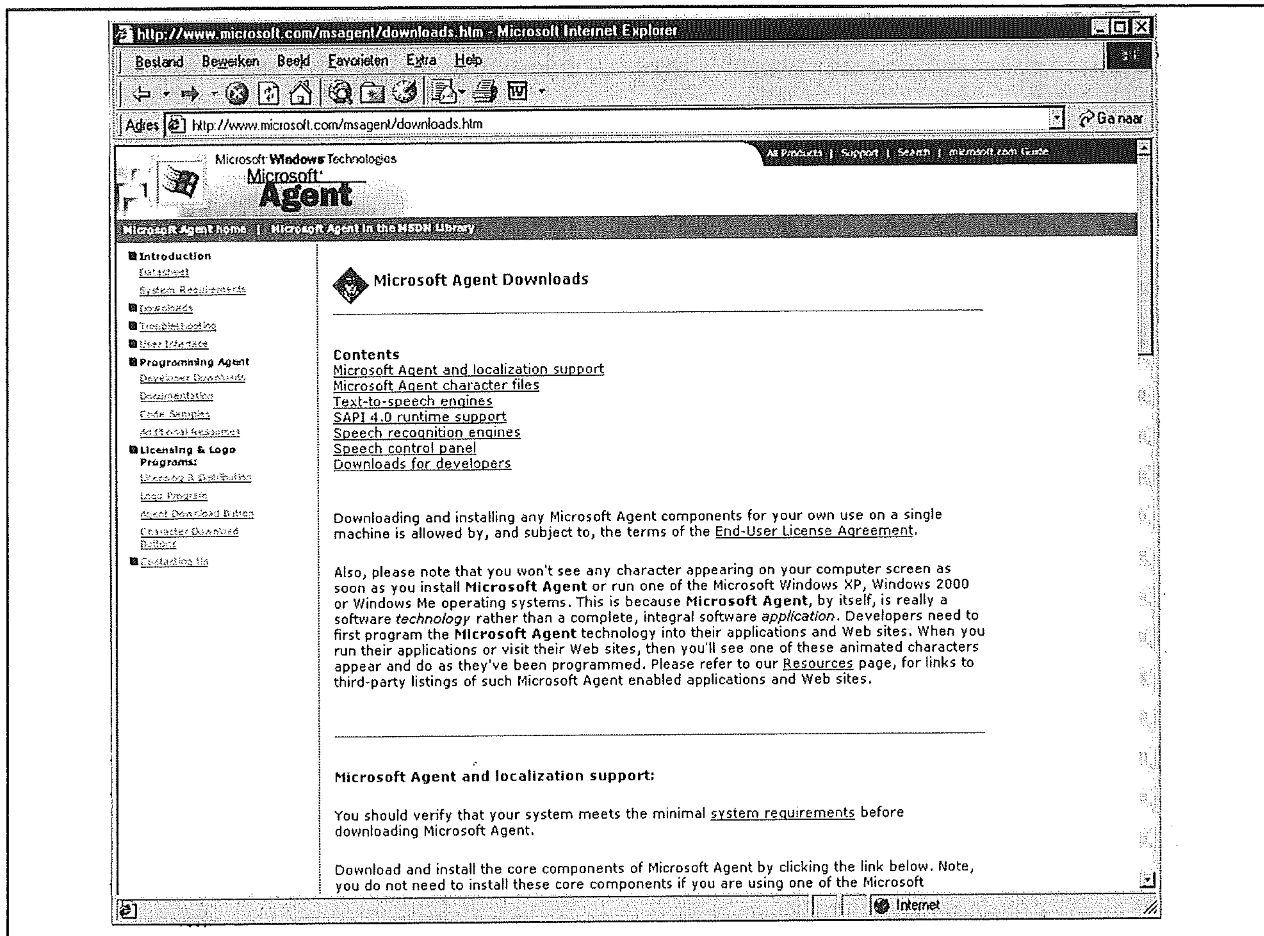
Er zijn heel wat Internetpagina's die behulpzaam zijn bij het downloaden van deze bestanden. Maar waarom niet meteen naar de bron van deze technologie, dus de Microsoft-site? Op de pagina www.microsoft.com/msagent/downloads.htm, zie figuur 3/19.15-5, vindt men een handige gids voor het downloaden van alle noodzakelijke bestanden.

De Microsoft Agent core components

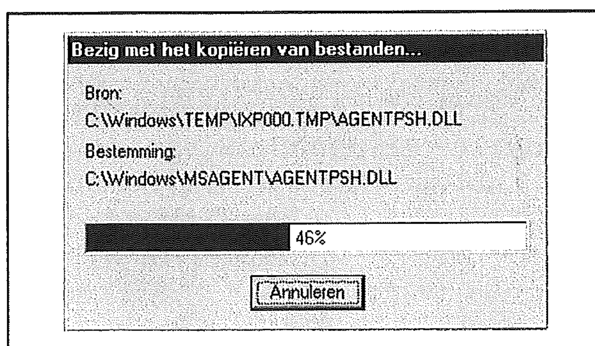
De basiscomponenten van de Agent-technologie zitten in het bestand MSAGENT.EXE (400 kB). Dit bestand maakt automatisch een directory C:\WINDOWS\MSAGENT of C:\WINNT\MSAGENT aan, waarin voornamelijk DLL's worden geïnstalleerd en het programma AGENTSVR.EXE. Deze Agent Server is de basis van de technologie.

Daarnaast zorgen deze DLL's ervoor dat Agent-applicaties de ingebouwde Agent Control en de Agent Provider kunnen uitvoeren.

19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-5: Van deze pagina uit de Microsoft-site kan men de noodzakelijke componenten downloaden.



Figuur 3/19.15-6: De belangrijkste onderdelen van de Agent-technologie worden geïnstalleerd: de Agent Server, Control en Provider.

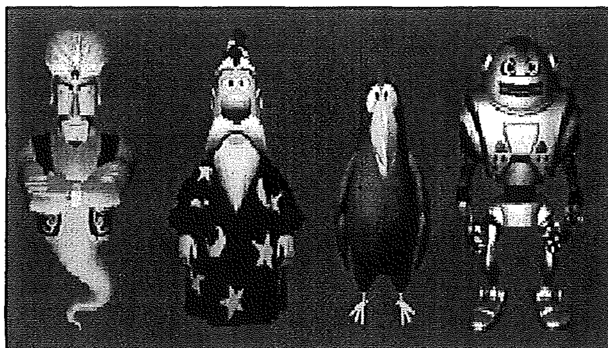
Dat zijn componenten die als interface dienen tussen de Agents die in het systeem geïnstalleerd worden (zie later) en de applicaties.

De Microsoft Agent Character-bestanden

Microsoft heeft zélf vier Agents in de aanbieding, zie figuur 3/19.15-7:

- “Robby the Robot”, een leuke robot met een omvang van 2,12 MB;
- “Merlin the Wizard”, een wijze tovenaar met een omvang van 1,78 MB;
- “Genie”, de geest uit de fles met een omvang van 1,55 MB;
- “Peedy the Parrot”, een eigenwijze papegaai met een omvang van 3,24 MB.

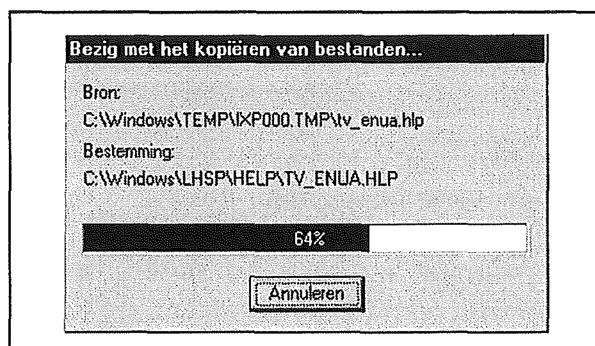
19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-7: De vier Agents van Microsoft.

Aanbevolen wordt in elk geval "Peedy" te downloaden. Met zijn prachtige animaties biedt deze ondertussen zeer populaire papegaai zeer uitgebreide, grappige en verrassende bewegingen en animaties. Men kan er tientallen minuten lang geamuseerd naar kijken. Waarmee meteen ook het verschil in omvang tussen de diverse Agents is verklaard. Hoe groter het bestand, hoe meer animaties in de Agent zijn verborgen en hoe meer men ermee kan doen.
De Agents zijn EXE-bestanden die automatisch een subdirectory CHARS in MS-AGENT aanmaken en daar één ACS-bestand in neerzetten. ACS is de afkorting van "Agent Character Set" en zo'n bestand bevat alle animaties en frames van een Agent. Er zijn Agents die meer dan tachtig animaties bevatten. Agent-applicaties gebruiken de ACS-bestanden om een bepaalde animatie uit de set te laden.

De Lernout & Hauspie TruVoice text-to-speech engine, Amerikaans-Engels
Op de genoemde pagina van Microsoft ziet men vervolgens het kopje "Text-to-speech"-engines. Men selecteert hier de Amerikaanse versie TV_ENUA.EXE. Een bestand van ongeveer 1 MB die de Agents van tien Amerikaanse stemmen voorziet. Windows werkt in de Agent-technologie samen met het Belgische bedrijf Lernout & Hauspie. Deze hebben engines ontwikkeld waarmee TXT-bestanden worden omgezet in spraak. Het genoemde bestand maakt automatisch een subdirectory LHSP aan in Windows en zet daar een groot aantal DLL's in.

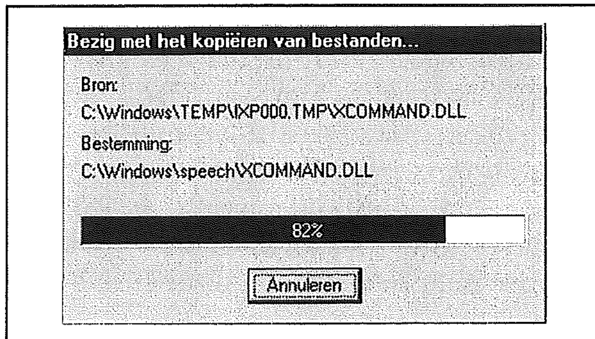


Figuur 3/19.15-8: De Lernout & Hauspie-bestanden worden volledig automatisch op het systeem geïnstalleerd.

De SAPI 4.0 runtime support
Dit bestand SPCHAPI.EXE van 844 kB is de spraak-API die de communicatie verzorgt tussen de spraakengines van Lernout & Hauspie en de Microsoft Agents. Ook dit bestand installeert zich volledig automatisch op het systeem en wel in de nieuwe subdirectory SPEECH van Windows, zie figuur 3/19.15-9.

Het Microsoft Speech Control Panel
Deze applicatie voegt een item "Speech" toe aan het configuratiescherm. Via dit item kan men alle geïnstalleerde spraakengines van Lernout & Hauspie configureren (zie later). Het te downloaden bestand SPCHCPL.EXE is 927 kB groot en kopieert een heleboel bestanden naar de SYSTEM-directory.

19.15 De Agent-technologie van Microsoft

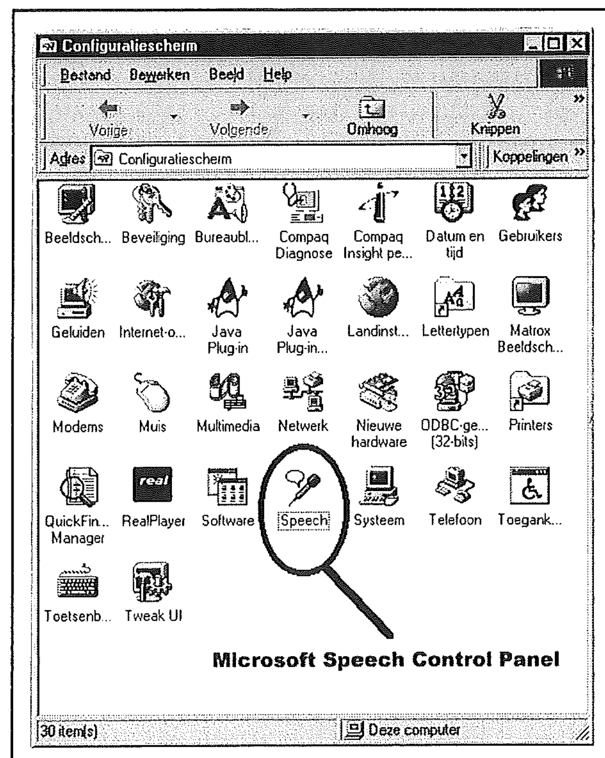


Figuur 3/19.15-9: Deze spraak-API zorgt voor de communicatie tussen de spraakengines van Lernout & Hauspie en de Agents.

In principe klaar, maar voor Nederlanders en Vlamingen nog nét niet

In principe kan men nu, na het opnieuw opstarten van het systeem, aan de slag met de Agent-technologie. Maar voor Nederlanders en Vlamingen zit er een addertje onder het gras. De geïnstalleerde spraakengine van Lernout & Hauspie leest TXT-bestanden voor met een Amerikaanse stem en die stem is niet in staat iets te bakken van tweeklinkers zoals in "doen". Het Belgische bedrijf heeft echter een heleboel andere TruVoice-engines ontworpen die rekening houden met de specifieke fonetische eigenschappen van een taal. Lernout&Hauspie vat deze samen met "TTS3000", afkorting van Text-To-Speech". Het is dus handig in elk geval ook de Nederlandse spraakengine van Lernout & Hauspie te downloaden. Daarnaast moet men echter voor elke andere taal dan Amerikaans-Engels ook nog eens een paar DLL's installeren. Complex? Ja, maar aan de hand van deze handleiding lukt het allemaal zonder problemen. Ga dus weer naar de pagina www.microsoft.com/msagent/downloads.htm en kies in het selectievakje "Select a language" voor "Dutch". Nu wordt het bestand AGTX0413.EXE (132 kB) gedownload dat niets anders doet dan een paar specifieke DDL's kopiëren naar de reeds bestaande MSAGENT-directory.

Vervolgens gaat men naar "Select a text-to-speech engine" en ook daar kiest men natuurlijk voor "Dutch". Het bestand LHTTSDUN.EXE van 2,7 MB voegt, ook alweer volledig automatisch, de bestanden van de Nederlandstalige spraakengine toe aan de reeds bestaande subdirectory LHSP.

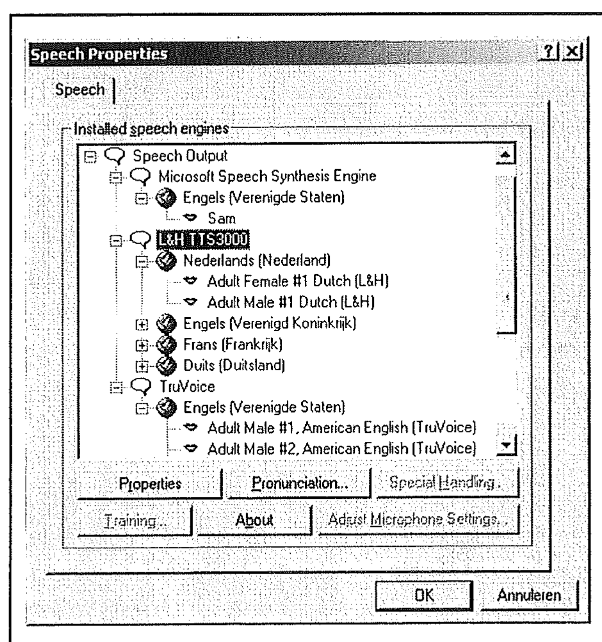


Figuur 3/19.15-10: Het bestand SPCHCPL.EXE maakt een nieuw controlepaneel aan in het configuratiescherm, namelijk "Speech".

Conclusie

Het Windows-systeem is nu klaar om de Agent-technologie van Microsoft met Nederlandstalige spraakuitgave te kunnen toepassen. In deze minimale configuratie zijn inmiddels een sappige 9.243 kB aan bestanden gedownload. Dat is, het zal dui-

delijk zijn, meteen de bottleneck van het systeem. Wil iemand Agents integreren in een Internetsite? Het is natuurlijk absurd te veronderstellen dat bezoekers bereid zijn ongeveer 10 MB te downloaden alvorens zij deze nieuwe Agent-pagina's kunnen bewonderen. Zolang niet alle Agent-componenten op alle PC's aanwezig zijn, is het niet verstandig Internetpagina's men Agents te ontwerpen.



Figuur 3/19.15-11: Het is gelukt! In het configuratiepaneel "Speech" blijkt duidelijk dat de Nederlandse en Amerikaans-Engelse stemmen van het Agent-systeem ter beschikking staan.

Het systeem opnieuw opstarten

Na het opnieuw opstarten van de PC wordt het configuratiescherm geopend en het paneel "Speech" aangeklikt. Als alles naar wens is verlopen, staan daar inderdaad onder het kopje "Speech Output" dat de Amerikaanse en Nederlandse spraakengines ter beschikking, zie figuur 3/19.15-11. De Nederlandstalige engine "L&H TTS3000" heeft twee stemmen. Een man, die Hans wordt genoemd, en een vrouw die door het leven gaat met de naam Karen. De Amerikaans-Engelse engine "TruVoice" beschikt over tien stemmen, acht mannen en twee vrouwen. Dit paneel bevat de drukknoppen "Properties", "Pronunciation", "Special Handling", "Training", "About" en "Adjust Microphone Setting". Allemaal zeer interessant, maar daar hebben we het nu nog even niet over. Alles op zijn tijd en aan de instellingen van de "Speech"-controlepaneel wordt een afzonderlijke paragraaf gewijd.

Andere talen

Het zal vooral onze Vlaamse lezers ten zeerste interesseren dat Lernout & Hauspie natuurlijk ook spraakengines voor Frans en Duits heeft ontwikkeld. Via de reeds vaker genoemde pagina www.microsoft.com/msagent/downloads.htm kan men deze downloaden. Het principe is steeds hetzelfde. Men selecteert eerst in het vakje "Select a language" het EXE-bestand waarmee de DLL's worden gedownload. Nadien gaat men naar het vakje "Select a text-to-speech engine" en selecteert daar dezelfde taal. Nadien voert men de twee EXE-bestanden uit en de componenten worden automatisch geïnstalleerd. Omdat het risico bestaat dat men op het laatst door de bomen het bos niet meer ziet, geven wij een overzicht van de namen van de bestanden voor de voornaamste Europese talen:

- Engels: LHTTSENG.EXE, geen DLL's noodzakelijk;
- Frans: LHTTSFRF.EXE en AGTX040C.EXE;
- Duits: LHTTSGED.EXE en AGTX0407.EXE;

- Italiaans: LHTTSITI.EXE en AGTX0410.EXE;
- Portugees: LHTTSPTB.EXE en AGTX0416.EXE;
- Spaans: LHTTSSPE.EXE en AGTX0C0A.EXE;
- Russisch: LHTTSRUR.EXE en AGTX0419.EXE.

In de meeste gevallen krijgt men er per taal een mannelijke en een vrouwelijke stem bij, zoals de Duitsers Stefan en Anna, de Fransen Pierre en Véronique en de Engelsen Peter en Carol.

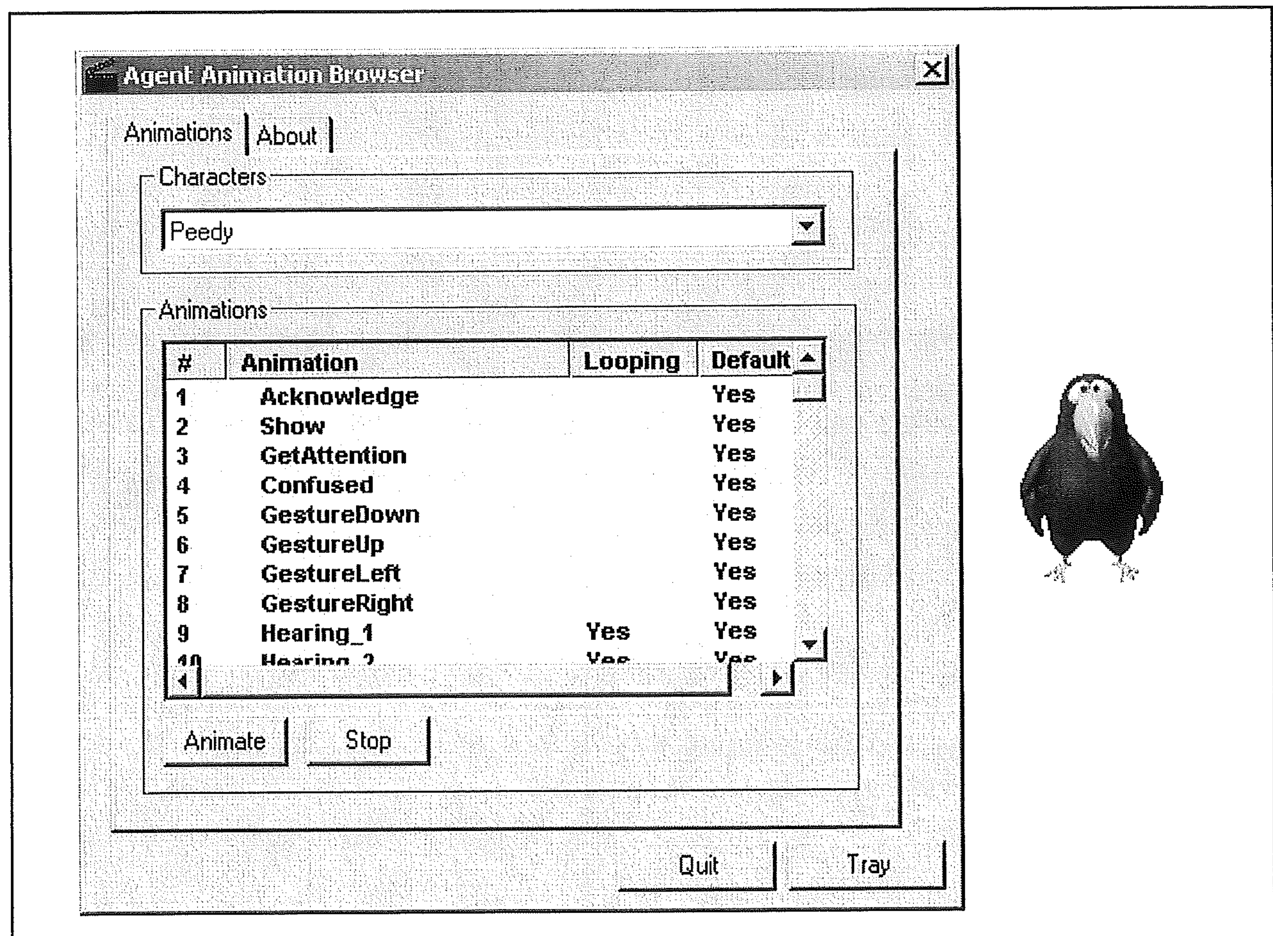
De eerste kennismaking

Klaar voor de start
Men heeft nu alle componenten geïnstalleerd die nodig zijn voor het werken met de Agent-technologie. Althans, wat betreft de animatie en de uitvoer van spraak; het derde kenmerk van deze technologie, het interpreteren van via een microfoon ingesproken commando's, laten we op dit moment voor wat het is. Men merkt echter nog niets van de aanwezigheid van de Agent-technologie, en dat klopt, want Agent is geen op zichzelf staande applicatie maar een uitbreiding van het Windows-besturingssysteem die wacht tot er een beroep op wordt gedaan. Om dat beroep te kunnen doen, is een heleboel programmeerkennis vereist. Laten we dus voor deze eerste kennismaking maar gebruik maken van programma's die kant en klaar van het Internet te downloaden zijn.

Animatie en spraak
De Agents kunnen bewegen en spreken. In de volgende drie paragrafen worden drie applicaties besproken, die deze twee aspecten van de Agents exploreren.

Agent Animation Browser
Met dit kleine programma, ontwikkeld door Softnik Technologies, kan men alle animaties die in een Agent verborgen zitten bestuderen. Als men namelijk meerdere Agents downloadt, zal men vaststellen dat de omvang van de bestanden nogal verschilt. Er worden Agents aangeboden die maar een paar tientallen kB's groot zijn. In principe ideaal voor gebruik op het Internet, maar vaak zijn de animaties zo beperkt, dat er weinig mee is aan te vangen. Dit programma is dus een uitstekend hulpmiddel om de kwaliteit van een Agent te beoordelen.
Het bestand AGANIMS.ZIP (130 kB) staat ter beschikking op softnik.com/products/aganim. Let op! Er staat dus géén www voor de URL! Na uitpakken kan men het bestand AGANIMS.EXE zonder meer opstarten; er worden geen DLL's en andere bestanden geïnstalleerd, zie figuur 3/19.15-12. In dit geval is Peedy tot onderwerp van de studie gemaakt. Selecteer in de lijst een animatienaam en klik op de knop "Animate". Wat blijkt? Dit meesterwerkje van Microsoft bevat niet minder dan 85 animaties, die overigens niet allemaal even overweldigend zijn. Maar kijk eens naar "Show", "Confused", "Hide", "Write", "Read" of "Suggest" en geniet van de Walt Disney-achtige animaties. Alle animaties hebben een unieke naam waarmee zij door programmeurs worden aangeroepen. Met de "Agent Animation Browser" vallen onwaardige Agents onmiddellijk door de mand. Kijk maar eens naar de minkukel van figuur 3/19.15-13, die weliswaar maar 40 kB groot is maar die maar één animatie bevat. Deze mislukte tijger is overigens een van de honderden Agents die gratis van het Internet kunnen worden gedownload. Maar daarover later meer.

19.15 De Agent-technologie van Microsoft

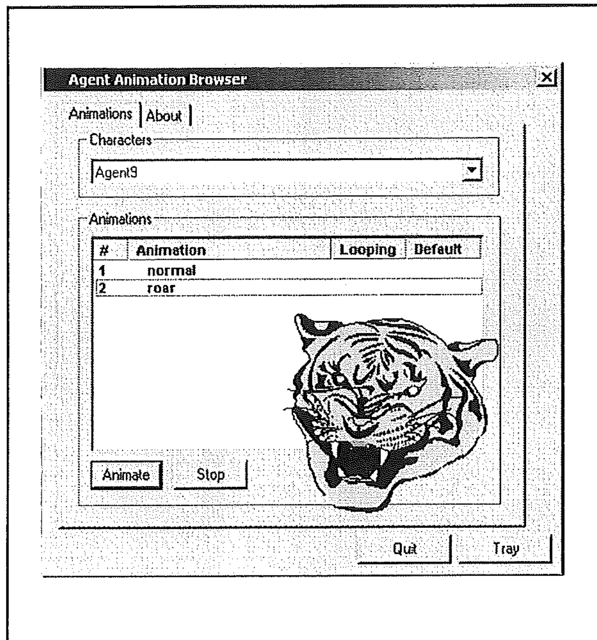


Figuur 3/19.15-12: Met dit kleine programma kan men alle animaties die in een Agent zitten bestuderen.

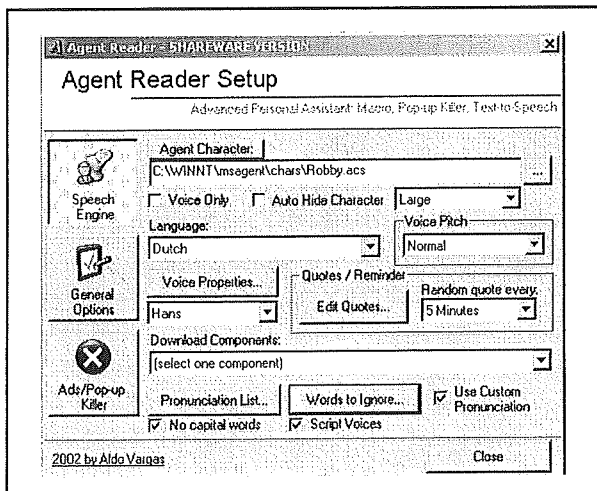
Aldo's Agent Reader

Nu de spraakuitvoer van de Agents. Er zijn tientallen programma's die de Agents laten spreken, maar een mooie instap is de Agent Reader van Aldo Vargas. Dit programma kan men downloaden van www.aldostools.com en heeft een omvang van 368 kB. Het is voorzien van een installer, zodat het installeren vanzelf gaat. In principe kan men met dit programma veel meer dan alleen Agents laten lezen, maar wij beperken ons hier tot deze functie. Na het opstarten van het programma nestelt de Agent Reader zich in de taakbalk. Door met de rechter muisknop op het pictogram te klikken, verschijnt een pop-up menu waarin men de optie "Options Setup" aanklikt. In het venster dat nu verschijnt, kan men de spraakuitgave volledig configureren, zie figuur 3/19.15-14. In het bovenste vakje "Agent Character" kiest men uiteraard de gewenste Agent, bijvoorbeeld Robby. In het vakje "Language" kan men de taal selecteren. Opgelet: het programma onderzoekt niet welke taalengines zijn geïnstalleerd, maar zet alle talen die door Lernout & Hauspie worden ondersteund in het lijstje. Kiest men een taal waarvan de engine niet is geïnstalleerd, dan gebeurt er vervolgens niets meer. het vakje "Voice Properties" selecteert men een van de stemmen, in dit geval Karen of Hans. De "Voice Pitch" laat men voorlopig maar op "Normal" staan.

19.15 De Agent-technologie van Microsoft

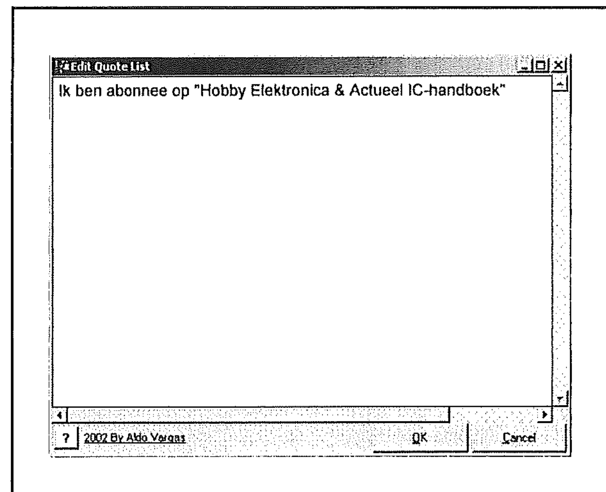


Figuur 3/19.15-13: Er zijn ook mislukte Agents te downloaden, zoals deze gehandicapte tijger die maar één animatie bevat.



Figuur 3/19.15-14: Het configuratiescherm van Aldo's Agent Reader.

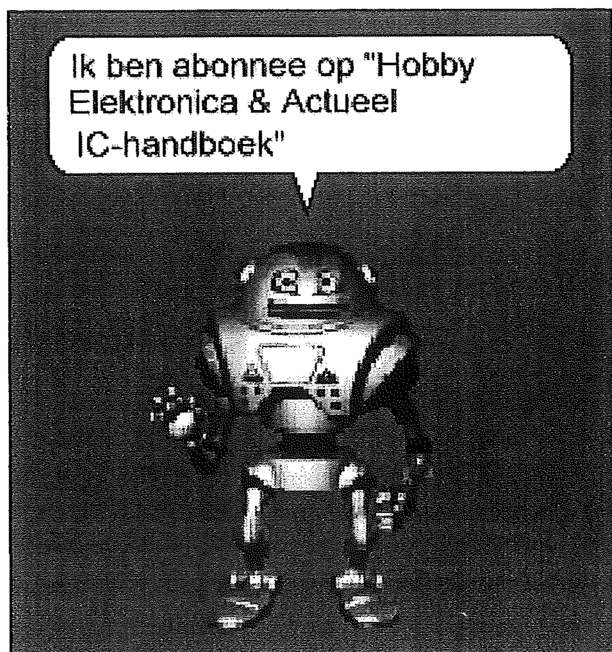
Op het moment dat de stem is ingesteld, verschijnt er een tekstballon boven de Agent die in duidelijk Nederlands "een, twee, drie" zegt. Het begin is dus veelbelovend. Met de utility van Aldo Vargas kan men de Agent bijvoorbeeld om de vijf minuten de tijd laten uitspreken. Men heeft de mogelijkheid om een korte samenvatting van de voornaamste systeemeigenschappen te beluisteren, zoals vrij geheugen, ruimte op de harde schijf enzovoort. Men kan ook "Quotes" invoeren, een lijst met korte zinnen die de Agent om de zoveel tijd in een willekeurige volgorde uitspreekt, zie figuur 3/19.15-15.



Figuur 3/19.15-15: In dit venster voert men de "Quotes" in, korte zinnen die de Agent in willekeurige volgorde en met gedefinieerde tijdsintervallen uitspreekt.

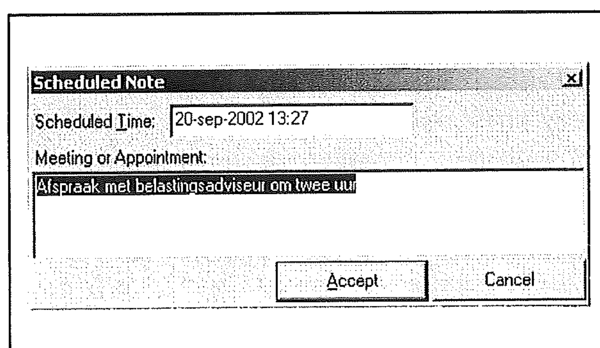
Tussen de bedrijven door is de Agent in rust. Het uitspreken van een quote gaat steeds gepaard met het uitvoeren van een van de animaties en een tekstballon waarin de tekst van de quote nog eens verschijnt, zie figuur 3/19.15-16. In het algemeen kan men stellen dat de uitspraak van het Nederlands verbazingwekkend duidelijk is. Tweeklanken zoals "oe" en "eu" worden goed geïnterpreteerd. Maar het blijft natuurlijk een mechanische stem, emotie in de zin leggen kan de Agent (nog) niet.

19.15 De Agent-technologie van Microsoft



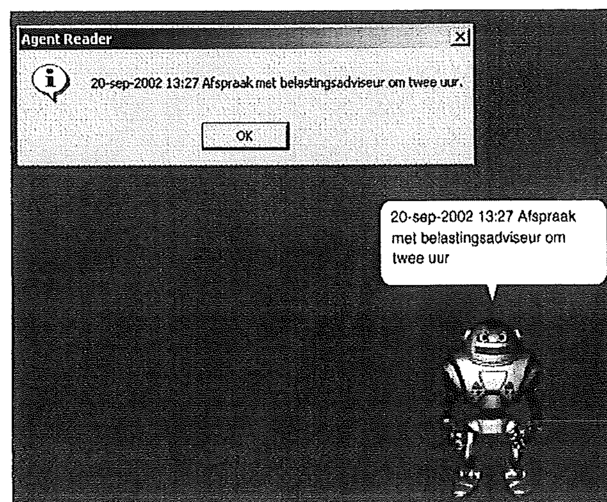
Figuur 3/19.15-16: Het uitspreken van een quote gaat gepaard met een tekstballon en een animatie.

Een handige functie van Aldo's Agent Reader is dat de Agent ook afspraken kan uitspreken. Men heeft daarvoor de beschikking over de gebruikelijke kalender en een tekstediter, zie figuur 3/19.15-17.



Figuur 3/19.15-17: Het invoeren van een nieuwe afspraak die nadien door de Agent wordt uitgesproken.

Men kan bijvoorbeeld invoeren dat de Agent er om half een aan herinnert dat om twee uur de belastingsadviseur voor de deur staat, zie figuur 3/19.15-18.



Figuur 3/19.15-18: De Agent spreekt op de afgesproken tijd de ingevoerde tekst uit.

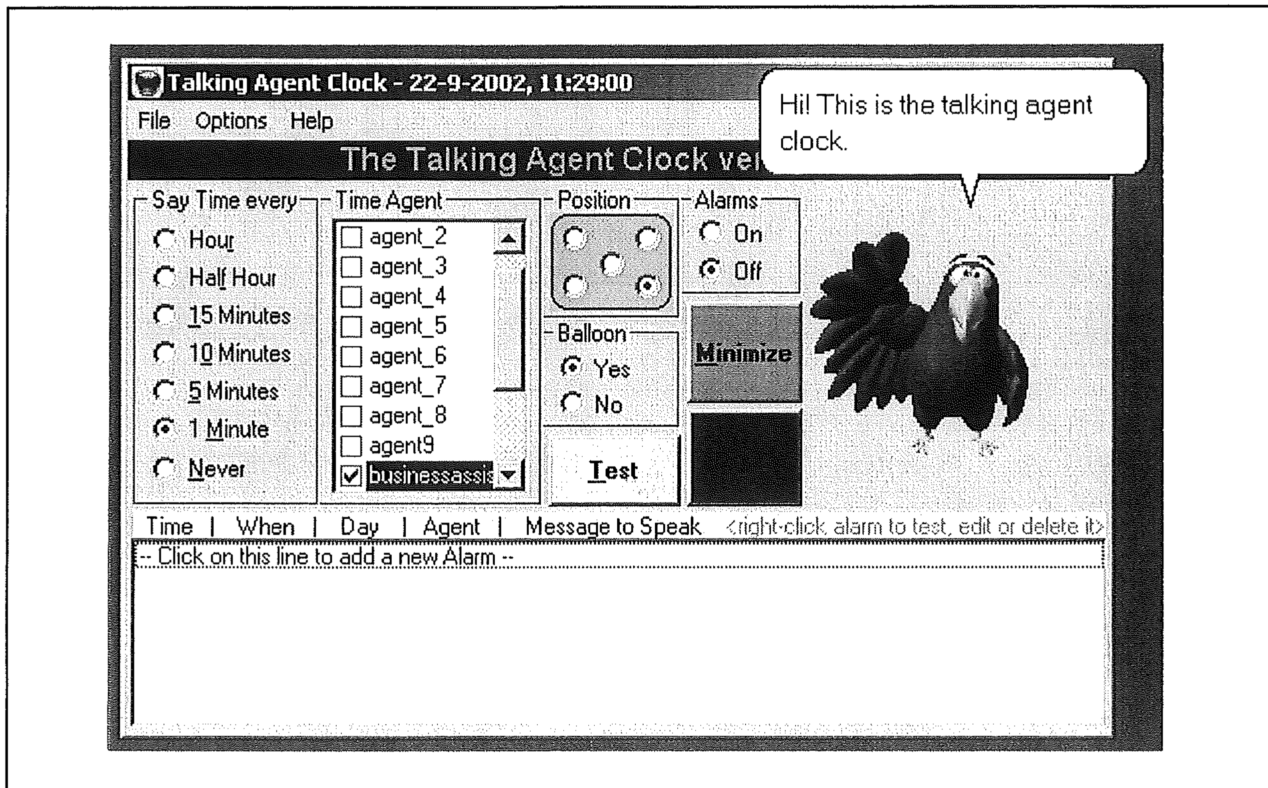
Talking Agent Clock

Dit programma vertoont veel gelijkenis met de Agent Reader. Echter, als men het helpmenu aanklikt, verschijnt er geen tekstvenster in beeld, maar komt een Agent tot leven die de helptekst in duidelijk Amerikaans uitspreekt, zie figuur 3/19.15-19.

Op dit thema zijn uiteraard talrijke variaties te verzinnen. Zo kan het bekende vraagtekentje dat in een paar zinnen uitleg geeft over een bepaalde optie in een venster, natuurlijk ook met een sprekende Agent worden uitgewerkt. Men kan het programma met een omvang van 1,8 MB downloaden van 207.233.110.35/tac. Uit de aard der zaak heeft het programma ook een sprekende klok. Men kan de tijduitgave instellen tussen elk uur tot elke minuut.

Daarnaast kan men echter ook alarmmeldingen invoeren, die het programma op de ingestelde tijd door een Agent laat uitspreken. Inderdaad, door een Agent, want men kan elke boodschap koppelen aan een andere Agent.

19.15 De Agent-technologie van Microsoft



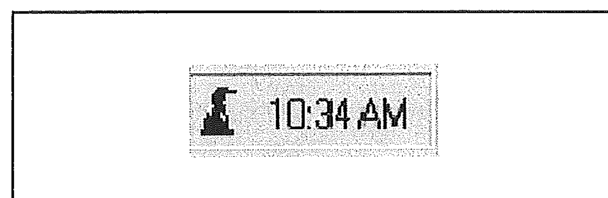
Figuur 3/19.15-19: Bij de Talking Agent Clock wordt de Agent-technologie toegepast om het helpbestand uit te spreken.

Zo kan men Robby inschakelen als herinnering dat elke woensdag de afvalcontainer wordt buitengezet. Helaas kan men bij dit programma geen stemmen instellen. De Talking Clock maakt gebruik van de standaardstemmen die door Microsoft worden meegeleverd. Het is dus verstandig de teksten in het Engels in te voeren, want het laten uitspreken van een Nederlandse tekst door de Amerikaanse stem klinkt vreselijk.

Conclusie
De Agent-technologie is dus meer dan alleen maar spielerei. De Aldo's Agent Reader is bovendien in feite maar een onbenullige applicatie. Men kan op dit moment meer dan 150 applicaties downloaden die op de een of andere manier een Agent integreren.

De Agents configureren

Inleiding
Als een applicatie een Agent aanroept, verschijnt deze niet alleen op het scherm, maar ook als pictogram in de taakbalk, zie figuur 3/19.15-20.



Figuur 3/19.15-20: De Agent, eenmaal aangeroepen door een applicatie, blijft actief in het geheugen van het systeem en meldt zijn aanwezigheid in de taakbalk.

19.15 De Agent-technologie van Microsoft

Klikken met de rechter muisknop op Agent of pictogram roept een pop-up menu op waarmee bepaalde eigenschappen van de Agent worden ingesteld. Microsoft wenst dat alle ontwerpers van Agent-applicaties een bepaalde standaardisatie aanhouden, zodat in elke applicatie de Agents op dezelfde manier worden geconfigureerd. Helaas wordt daar maar zelden rekening mee gehouden. Bij sommige applicaties geeft het pop-up menu alleen de keuze tussen de Agent verbergen of activeren, zie figuur 3/19.15-21. Let wel, de Agent blijft na het verbergen wél actief en kan op dezelfde manier weer op het scherm worden gebracht.



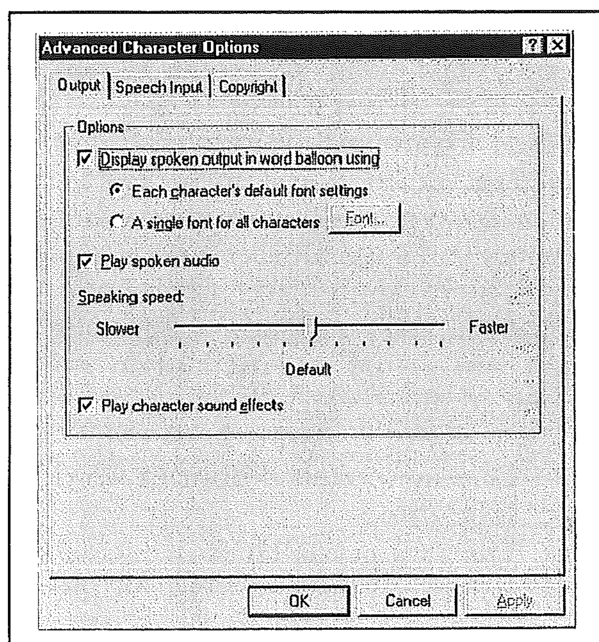
Figuur 3/19.15-21: Het pop-up menu van een Agent moet minimaal de optie bevatten om de animatie van het scherm te verbannen.

De configuratie-opties van Microsoft

Als de Agent-applicatie de richtlijnen van Microsoft opvolgt, ziet men in het pop-up menu de optie "Voice Command" of in het Nederlands "Venster gesproken opdrachten openen". Met deze optie kan men de mogelijkheden van de Agent-technologie instellen om via een microfoon ingesproken commando's te herkennen. Noteer dat dit menu dan ook niet met de muis of het toetsenbord is te benaderen, maar alleen via spraak. Op dit moment laten wij de mogelijkheden van deze "Voice Command Window" onbesproken. Het venster opent immers niet als de extra software van Microsoft voor het herkennen van spraakinvoer niet is geïnstalleerd. Interessanter is de menu-optie "Advanced Character Options". Hiermee kan men een aantal eigenschappen van de Agent configureren.

- Tab "Output", figuur 3/19.15-22
  - Display spoken output in word balloon using
    Hiermee bepaalt men of er een tekstballon wordt weergegeven als een Agent spreekt, dit in de veronderstelling dat de Agent een tekstballon voor uitvoer ondersteunt.
  - Each character default font settings
    Hiermee stelt men de individuele lettertype-instellingen in de tekstballon van een Agent in.
  - A single font for all characters
    Hiermee geeft men de tekstballonnen van de Agents weer met de opgegeven lettertype-instellingen.
  - Play spoken audio
    Hiermee bepaalt men of men gesproken uitvoer wenst of alleen boodschappen in de tekstballonnen.
  - Speaking speed
    Hiermee stelt men de relatieve snelheid in van de gesproken uitvoer voor alle Agents.
- Tab "Speech Input", figuur 3/19.15-23
  - Characters listen for input
    Hiermee bepaalt men of gesproken invoer voor de Agents wordt ingeschakeld. Deze optie is alleen van toepassing op Agents die gesproken invoer ondersteunen en als de spraakherkenningssoftware is geïnstalleerd.

19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-22: In de tab "Output" stelt men de tekst- en ballonuitvoer van de Agent in.

- When you press the ... key
  Hiermee stelt men de toets of toetsencombinatie in die men moet indrukken om de Agent te laten luisteren naar een stem. Dit noemt men "de luistertoets".
- And continu for ...
  Hiermee bepaalt men of het hulpvenster "Luistertips" wordt weergegeven als men op de luistertoets drukt.
- Play tone when you can speak
  Hiermee bepaalt men of er een attentiegeluid moet klinken om aan te geven wanneer men kan spreken.

De "Character Properties"

Met deze optie in het rechter muisknop menu kan men de zogenaamde "default Agent" selecteren. Applicaties kunnen elk afzonderlijk een Agent laden. Er bestaan echter ook applicaties die gebruik maken van een reeds geactiveerde Agent, die dus aanwezig is in het geheugen van het systeem. Dit wordt de "default Agent" genoemd. Via zo'n applicatie kan men dan geen Agent selecteren. Via dit venster kan men dat echter wél. Met de toetsen "Back" en "Next" kan men door de op de harde schijf aanwezige Agents scrollen en er een als default Agent uitkiezen, zie figuur 3/19.15-24.

De configuratie van de taal-engines van Lernout & Hauspie

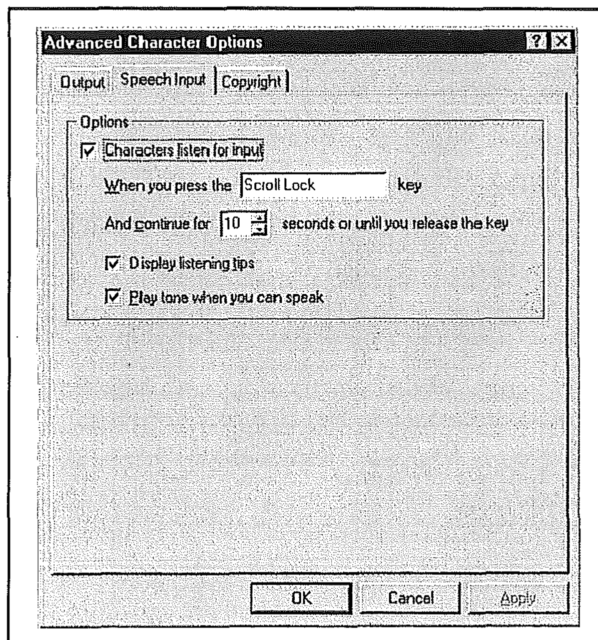
De taalengines van Lernout & Hauspie treft men aan onder het pictogram "Speech" van het configuratiescherm. Ook hier kan men aan de slag met het finetunen van het Agent-systeem. Als men op de entry "L&H 3000" klikt, verschijnen de knoppen "Properties", "Pronunciation" en "About".

Properties

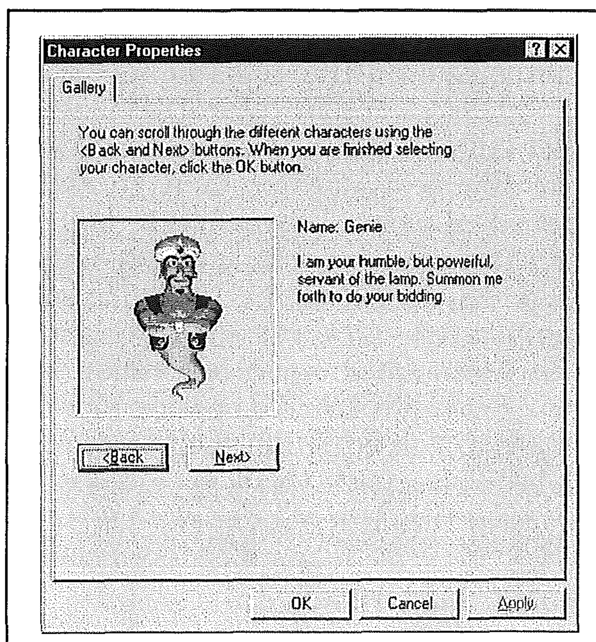
In het "Properties"-scherm dat in het Nederlands "Algemene parameters instellen" heet (zie figuur 3/19.15-25) kan men de eigenschappen van de stemmen instellen.

- Stemnaam
  Selecteer de gewenste spreker, dus vrouwelijk of mannelijk.
- Stemkenmerken
  Selecteer een stem uit de lijst. Men kan kiezen uit "Variatie 1" (standaardstem) tot "Variatie 9". Elke variatie heeft vaste waarden voor de toonhoogte en het spreektempo. Men kan ook een nieuwe stem maken door de variatie "Aangepast" te kiezen. De schuifregelaar voor de toonhoogte en het spreektempo kan men alleen gebruiken als de optie "Aangepast" is geselecteerd.
- Hoogte
  Met deze schuifregelaar stelt men de toonhoogte van de stem in.

19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-23: Met de tab "Speech Input" bepaalt men de opties om de Agents te laten luisteren naar een stem.



Figuur 3/19.15-24: Met dit venster kan men een van de aanwezige Agents selecteren als default.

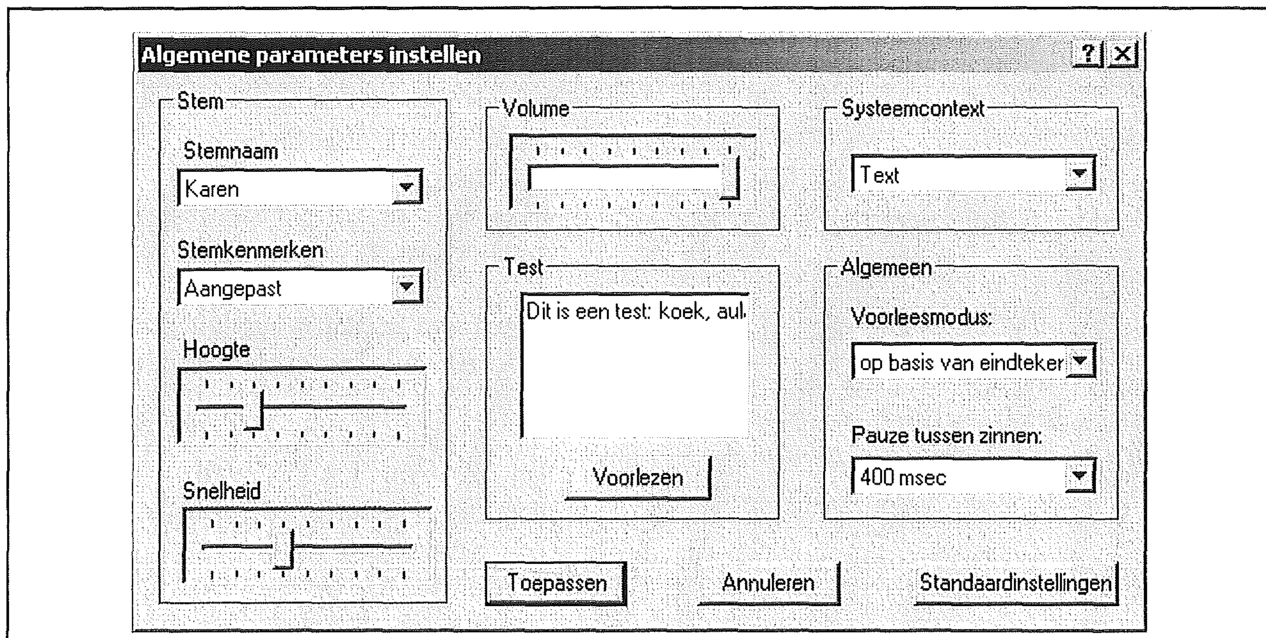
Aangezien de stemvariaties 1 tot 9 een vaste toonhoogte hebben, kan men de schuifregelaar alleen voor een aangepaste stem gebruiken.

- Snelheid
  Met deze schuifregelaar stelt men het spreektempo van de stem in.
- Volume
  Met deze schuifregelaar stelt men het volume van de geselecteerde stem in.
- Test
  Typ een zin in het testvak en klik op de knop "Voorlezen" om de tekst op basis van de geselecteerde parameters te laten voorlezen.
- Systeemcontext
  Selecteer de systeemcontext: kies "Text" als men standaardtekst wilt laten voorlezen; kies "EMail" als men e-mail berichten wilt laten voorlezen.
- Voorleesmodus
  Selecteer de voorleesmodus. Kies letter-voor-letter, woord-voor-woord, zin-voor-zin of op basis van eindteken. De ingevoerde tekst wordt standaard zin voor zin voorgelezen.
- Pauze tussen zinnen
  Geef de lengte op van de pauze tussen letters, woorden of zinnen, afhankelijk van de voorleesmodus. Men kan een pauze van 0 tot 1800 milliseconden instellen.

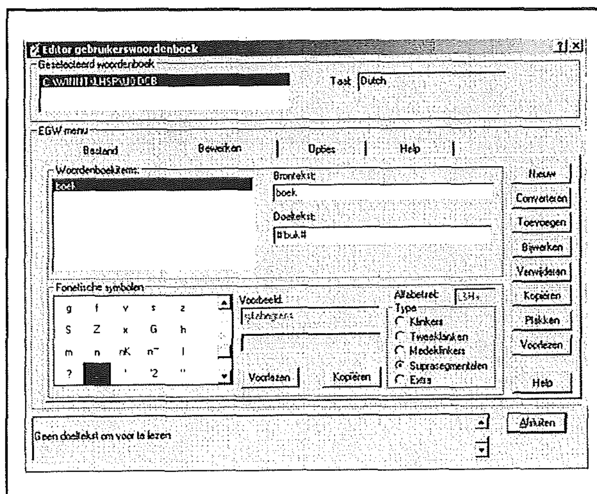
Pronunciation

Deze knop geeft toegang tot de "L&H-editor voor gebruikerswoordenboeken (EGW)", zie figuur 3/19.15-26. Nu komt het échte finetunen aan de beurt. Dit is een toepassing die is ontworpen voor het maken en bewerken van gebruikerswoordenboeken. Men kan met behulp van gebruikerswoordenboeken de uitvoer van het tekst-naar-spraaksysteem aanpassen.

19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-25: In dit venster "Algemene parameters instellen" kan men de eigenschappen van de twee Nederlandse stemmen Hans en Karen configureren.



Figuur 3/19.15-26: Via het tabblad "Bestand "van de "L&H-editor voor gebruikerswoordenboeken" kan men een eigen gebruikerswoordenboek aanmaken.

Men kan in deze woordenboeken een speciale uitspraak opgeven voor bepaalde woorden of tekenreeksen, bijvoorbeeld afkortingen, waarvan het standaardwoordenboek van L&H de uitspraak een beetje verhaspelt. Gebruikerswoordenboeken kunnen zowel orthografische (gewoon alfabet) als fonetische informatie bevatten. Het tabblad "Bestand" van "EGW" bevat alle functies waarmee men gebruikerswoordenboeken kan beheren. Men kan de naam van het gebruikerswoordenboek instellen, de taal van het gebruikerswoordenboek en een korte beschrijving invoeren. Het tabblad geeft bovendien informatie over het aantal items dat men heeft ingevoerd en de datum en tijd waarop het gebruikerswoordenboek is gemaakt. Het standaardwoordenboek van L&H heet U0.DCB en men kan dit niet openen en dus ook niet wijzigen.

In het tabblad "Opties" kan men het fonetische alfabet selecteren en de spraakparameters voor het tekst-naar-spraaksysteem opgeven. Er worden drie alfabetset's ondersteund:

(wordt vervolgd)

19.15 De Agent-technologie van Microsoft

- het gewone alfabet;
- het L&H+ fonetische alfabet;
- het internationale fonetische alfabet IPA.

Voor elk van de fonetische alfabetten biedt EGW een foneemset. Hierdoor wordt het invoeren van de doeltekst eenvoudiger als men het fonetische alfabet niet kent.

Het tabblad "Bewerken" is uiteraard het belangrijkste. Hier kan men woorden, waarvan de standaarduitspraak onduidelijk of niet mooi is, aanpassen. Zo spreekt de Nederlandse stem van L&H het woord "boek" een beetje vreemd uit. De tweeklank "oe" wordt in feite te kort aangehouden. In het vak "Brontekst" voert men dus het woord "boek" in. In het vakje "Doeltekst" verschijnt nu de fonetische samenstelling van dit woord zoals het in het standaardwoordenboek van L&H is opgenomen. In dit geval is als alfabetset "L&H+" gekozen. In het vak "Fonetische symbolen" verschijnt een overzicht van alle fonetische symbolen die in die alfabetset ter beschikking staan. Handig, want de meeste mensen hebben waarschijnlijk geen enkele kennis van fonetische alfabetten. Om het nog gemakkelijker te maken, verschijnt in "Type" het soort foneem:

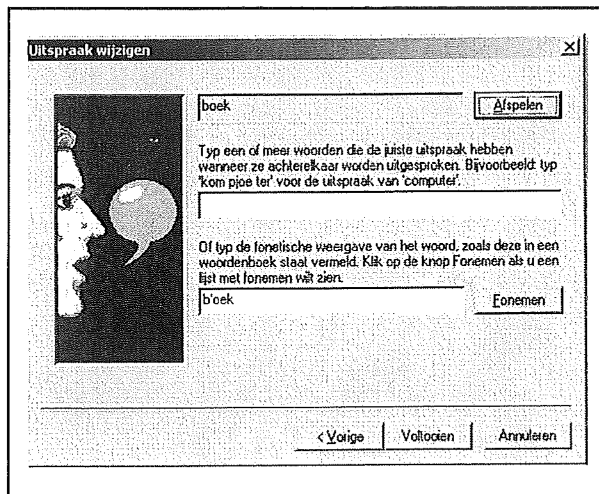
- een klinker;
- een tweeklank;
- een medeklinker;
- een suprasegment;
- een extra code.

De foneemset "suprasegment talen" bevat de symbolen die worden gebruikt om de syllabegrens weer te geven, evenals de primaire en secundaire woordklemtoon. De foneemset "extra" bevat de symbolen voor zinsaccent, stilte, woordgrens en leestekens.

Vervolgens moet een foneem uit het lijstje worden geselecteerd dat aan de uitspraak van het woord "boek" wordt toegevoegd. Men klikt het foneem aan en door op de knop "Voorlezen" te drukken, wordt een woord uitgesproken waar dit foneem in voorkomt. Op deze manier wordt op zoek gegaan naar het foneem dat ontbreekt in de uitspraak van het woord "boek". Gevonden? Klik dan op de knop "Kopiëren" en het foneem wordt toegevoegd in het vakje "Doeltekst". Rechts in beeld ziet men een heleboel knoppen die de volgende functies hebben.

- Nieuw
  Wist de vakken "Brontekst" en "Doeltekst", zodat er een nieuw woordenboekitem kan worden ingevoerd.
- Converteren
  Converteert het woord in het vak "Brontekst" naar zijn fonetische transcriptie. Het resultaat verschijnt in het vak "Doeltekst". Het alfabet dat wordt gebruikt voor de doeltekst is afhankelijk van het fonetische alfabet dat is geselecteerd op het tabblad "Opties".
- Toevoegen
  Voegt een nieuw woordenboekitem aan het gebruikerswoordenboek toe.
- Bijwerken
  Werkt de doeltekst van een item in het gebruikerswoordenboek bij.
- Verwijderen
  Verwijdert een of meer items uit het gebruikerswoordenboek.
- Kopiëren
  Kopieert een of meer items uit het gebruikerswoordenboek naar het klembord.
- Plakken
  Plakt alle gekopieerde woordenboekitems in het woordenboek.
- Voorlezen
  Klik op deze knop om het geselecteerde woordenboekitem te laten voorlezen door het tekst-naar-spraaksysteem.

19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-27: Ook de ene standaardstem die door Microsoft wordt geïnstalleerd, kan helemaal aan de eigen wensen worden aangepast.

De "Microsoft Speech Synthesis Engine" configureren

Bij de installatie van de "Microsoft Agent core components" wordt standaard één Amerikaanse stem geïnstalleerd. Ook die kan men volledig configureren. Ga in de "Speech"-entry van het configuratiescherm naar "Microsoft Speech Synthesis Engine" en klik weer op de knop "Pronunciation".

In het venster dat nu verschijnt, voert men een woord in waarvan de uitspraak niet bevalt, als voorbeeld maar weer eens "boek". In het volgende scherm staat dat woord en kan men op een primitieve manier de uitspraak van het woord samenstellen door lettergrepen in te voeren. Als voorbeeld wordt gesteld dat de uitspraak van het woord "computer" wordt verbeterd door dit in te voeren als "kom pjou ter". Maar helaas werkt dit in de meeste gevallen niet. Gelukkig kan men in het onderste kader ook de fonetische spelling definiëren. Hier hebben wij de uitspraak verbeterd door de fonetische code "'" voor de tweeklank "oe" op te nemen. Dit is het zogenaamde "primary stress"-symbool dat de nadruk in een woord regelt, zie figuur 3/19.15-27.

Microsofts "TruVoice" configureren

Door het installeren van het bestand TV_ENUA.EXE, de "Lernout&Hauspie TruVoice text-to-speech engine, Amerikaans-Engels", zijn er tien Amerikaanse stemmen aan het systeem toegevoegd. Deze stemmen zijn ontwikkeld door Lernout & Hauspie en door Microsoft aangekocht.

Ook die stemmen kan men weer op de nu reeds bekende manier configureren. Kies in het "Speech"-paneel voor "TruVoice" en klik weer op de knop "Pronunciation", zie figuur 3/19.15-28.

Ook nu is weer het woord "boek" als voorbeeld genomen. In het kader "Sounds Like" kan men een ander woord invoegen waarvan de uitspraak lijkt op deze van "boek".

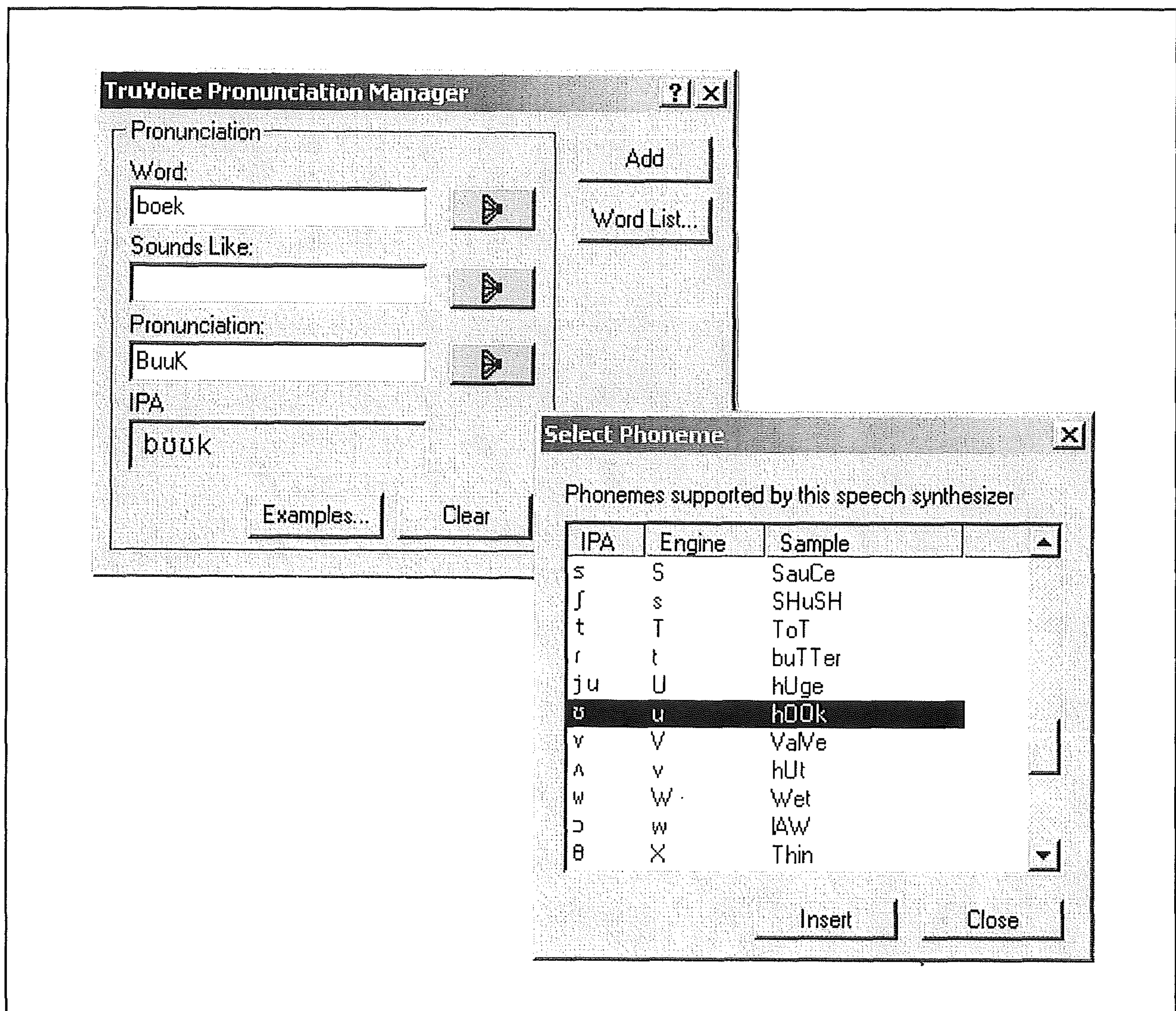
In het onderste kader kan men de uitspraak fonetisch instellen. Ook nu is weer een lijstje met fonemen aanwezig ("Word List") waarmee de uitspraak fonetisch wordt samengesteld. Als extra service geef het programma geheel gratis de codering weer in de internationale fonemencode IPA.

Eigen Agents verzamelen

Inleiding

Als alles volgens de richtlijnen is uitgevoerd, krijgt men nu toegang tot de vier Agents van Microsoft. Prachtige karakters, maar eerlijk gezegd worden deze zo vaak toegepast dat ze nogal afgelikt zijn.

19.15 De Agent-technologie van Microsoft

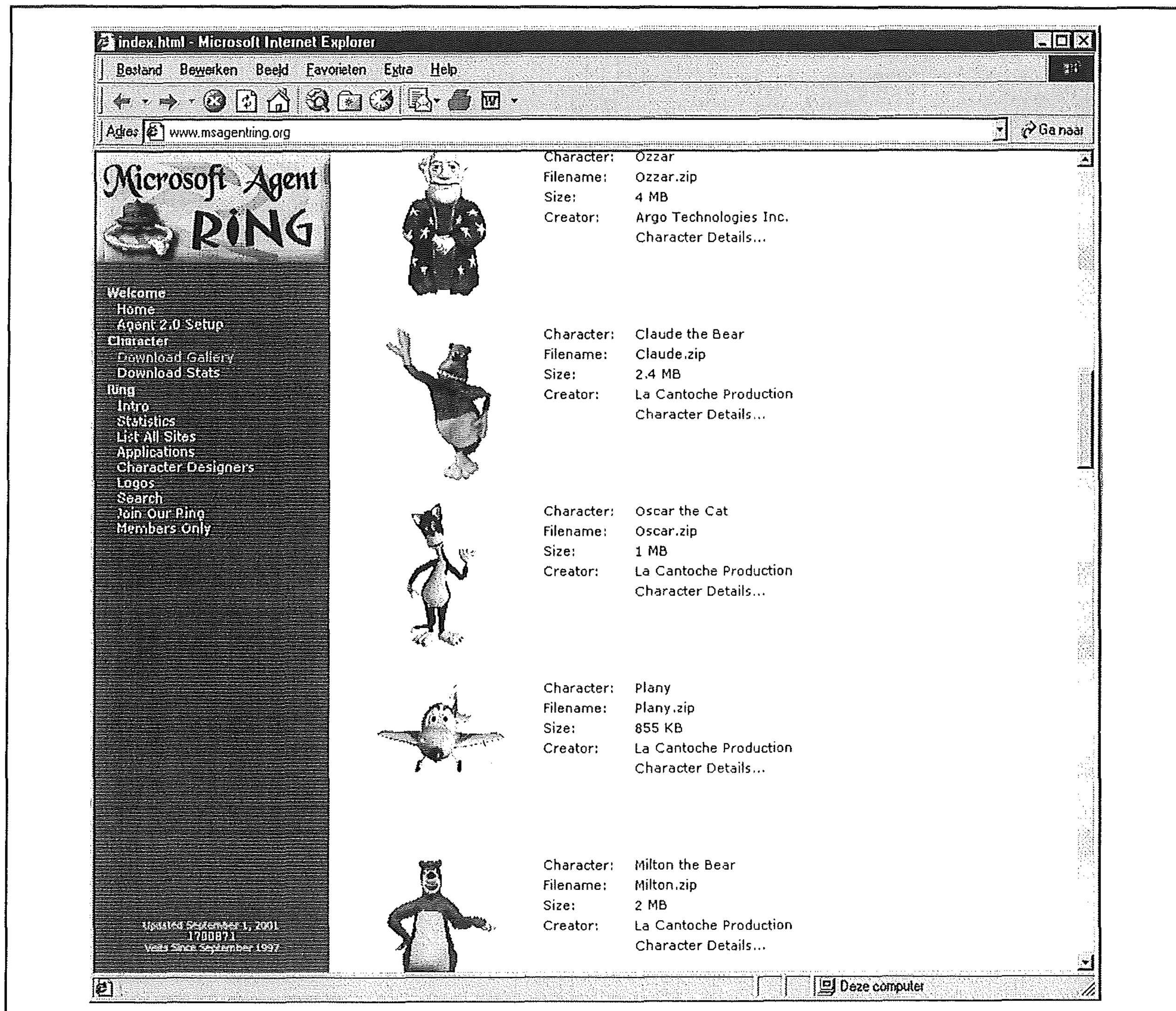


Figuur 3/19.15-28: Ook de uitspraak van de tien stemmen die het bestand TV_ENUA.EXE aan het systeem heeft toegevoegd, kan men configureren.

Wil men zélf commercieel aan de slag met de Agent-technologie, dan is het van belang dat een Agent wordt gebruikt die nog niet vaak wordt toegepast. Hoe komt men aan Agents? Daar bestaan vier bronnen voor:
- gratis Agents downloaden van het Internet;
- een Agent via het Internet kopen;
- een geheel eigen Agent laten ontwerpen;
- zélf een Agent ontwerpen.

Gratis Agents downloaden
Een goede start is www.msagentring.org, zie figuur 3/19.15-29, een site-ring waarbij tientallen aanbieders, die iets te melden hebben op het gebied van de Agent-technologie, zijn aangesloten. Via de optie "Download Gallery" komt men op een pagina waarvan men tientallen Agents gratis kan downloaden.

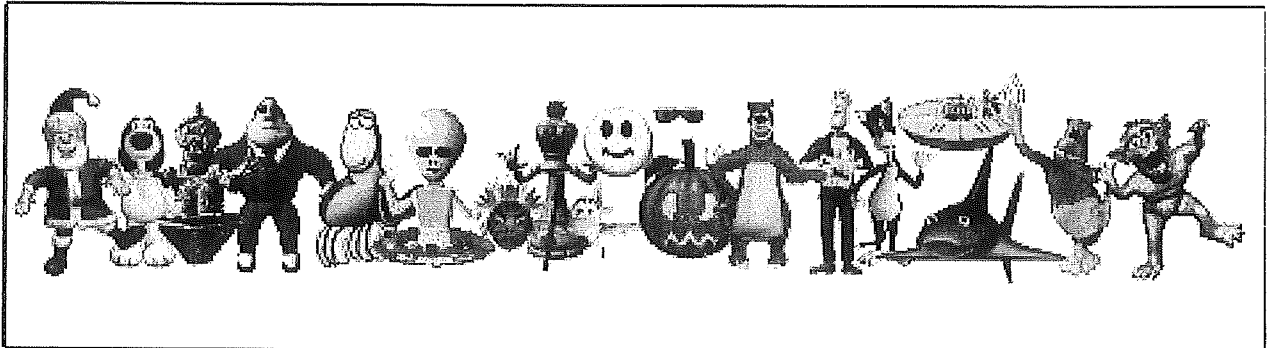
19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-29: Op de site van de Microsoft Agent Ring staat deze pagina, waarvan men tientallen gratis Agents kan downloaden.

Men moet er wel rekening mee houden dat er heel wat minderwaardige Agents in omloop zijn. Niet te verbazen, want het ontwerpen van een goed geanimeerde Agent vereist uiteraard heel wat grafisch vakmanschap.
Alleen een begenadigde animator is immers in staat vloeiend bewegende animaties te ontwerpen. Een eerste indicatie voor de kwaliteit van een Agent krijgt men door naar de bestandsgrootte te kijken. Als die minder dan 1 MB is, kan men er van uitgaan dat er weinig aandacht is besteed aan de animaties. Een tweede test is uiteraard gebruik te maken van het reeds besproken programma “Agent Animation Browser”. Maar, ondanks dit kaf onder het koren staan er werkelijk schitterende Agent gratis te beschikking, zie figuur 3/19.15-30.

19.15 De Agent-technologie van Microsoft

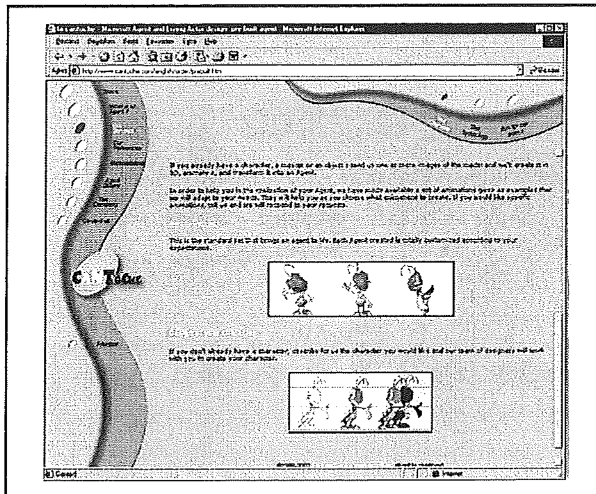


Figuur 3/19.15-30: Een twintigtal Agents die men kan downloaden uit de Microsoft Agent Ring.

e-Clips - Developers of outstanding Microsoft® Agent characters - Microsoft Internet Explorer
Bestand Bewerken Beeld Favorieten Extra Help
Adres http://www.e-clips.com.au/index.cfm?type=dsp_products&scid=16 Ga naar
E-CLIPS intelligent agent technologies
SEARC GO HOME | LOGIN | SHOPPING CART
Software > Characters > Animal
Sort by: Natural | Rating | Latest |
Desktop Wizard
IMP
Characters
Corporate Services
News
About Us
Contact e-Clips
E-Mail
go!
Squidge
Squidge is an ugly little guy but not as vulgar as WartNose. He is probably not the smartest guy you ever met either, (anyone who hits themselves on the head with a bone can't be smart.
Buy Download
$14.95
File Size: 3.3 Mb
Click Here For More
Zak
Zak is a major improvement on the mutt. He is the kind of dog that will always be there for you. He is playful, happy and willing to be your best friend. With approximately 60 actions he is so nice, you could not be without him.
Buy Download
$14.95
File Size: 4.5 Mb
Click Here For More
4 Hierarchical Menu Trees Created
Internet

Figuur 3/19.15-31: Een bedrijf dat goedkope Agents te koop aanbiedt is E-clips.

19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-32: Bij het Franse bedrijf Cantoche kan men een privé-Agent laten ontwerpen.

Een Agent kopen

Diverse bedrijven hebben zich gespecialiseerd in het ontwerpen van Agents. Soms worden deze Agents algemeen te koop aangeboden, wat wel als gevolg heeft dat men nog niet verzekerd is van een unieke Agent. Maar in elk geval zullen dit soort Agents natuurlijk veel minder vaak worden aangetroffen dan de gratis te downloaden exemplaren. Een bedrijf waar men heel snel via het Internet terecht kan voor de aanschaf van een Agents is E-clips, te vinden op www.eclips.com.au. Het systeem is te vergelijken met het aanschaffen van software via het Internet. Men selecteert een Agent en kan via een creditcard betalen. Nadien kan men het bestand downloaden. Het bedrijf biedt tientallen Agents te koop aan voor een gemiddelde prijs van US$ 15,00, zie figuur 3/19.15-31.

Een geheel eigen Agent laten ontwerpen

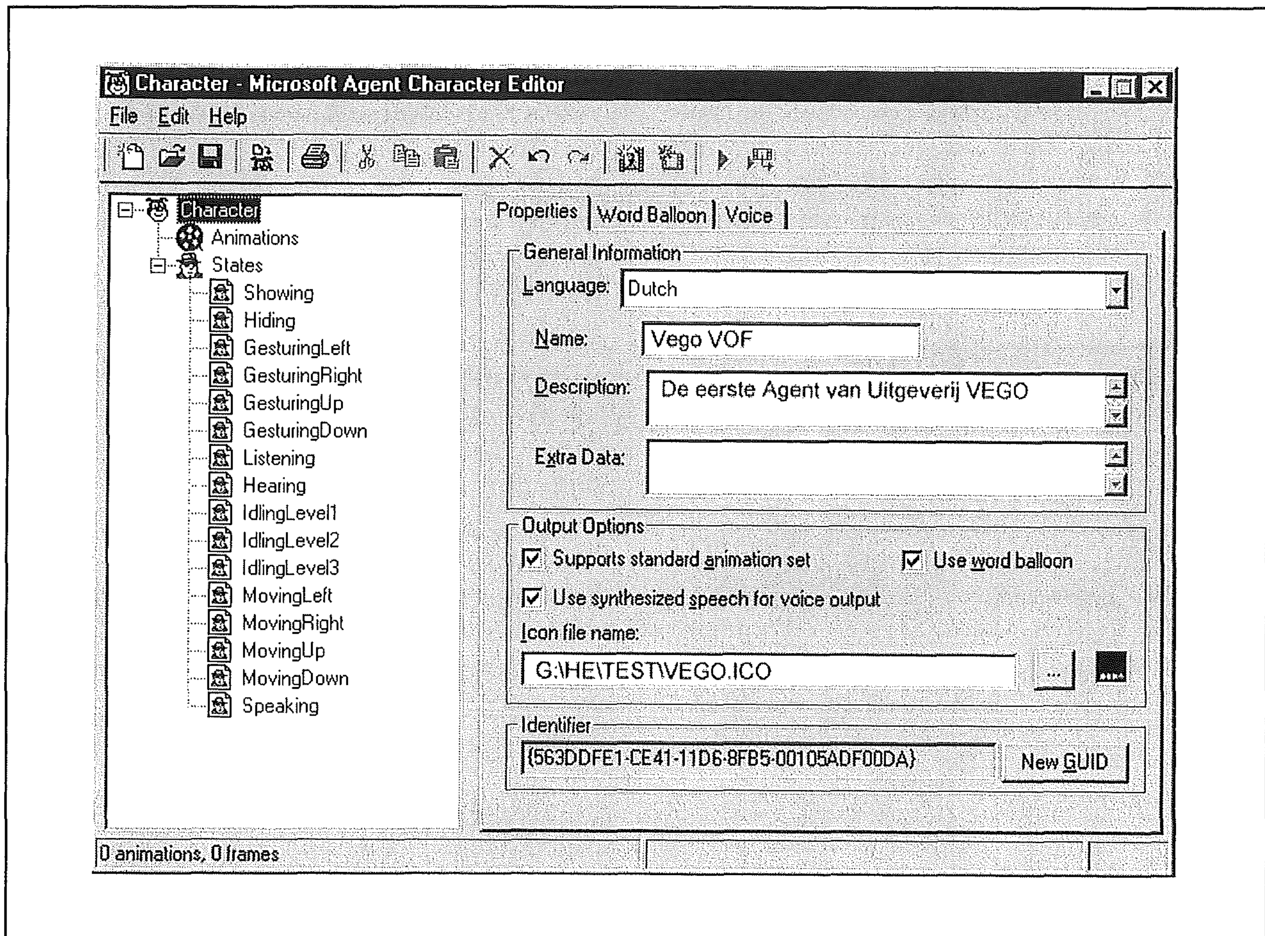
Het Franse bedrijf Cantoche (www.cantoche.com), zie figuur 3/19.15-32, heeft zich gespecialiseerd in het maken van elektronische animaties. Uiteraard hebben zij zich ook op de Agent-technologie gestort. Zij bieden de mogelijkheid een unieke Agent te ontwerpen. Wat dat grapje moet kosten? Daar wordt sereen over gezwegen, wel kan men een prijsofferte per e-mail aanvragen. Omdat de jongens achter Cantoche natuurlijk ook wel beseffen dat het laten ontwerpen van een unieke Agent voor de meeste bedrijven onbetaalbaar is, bieden zij een aantal tussenoplossingen aan. Zo kan men een van de door dit bedrijf reeds ontworpen Agents laten aanpassen aan de eigen wensen. Men kan er bijvoorbeeld een bedrijfslogo in laten integreren, men kan kleuren laten wijzigen en animaties uit het bestand laten verwijderen.

Zélf een Agent ontwerpen

Daar is natuurlijk speciale software voor nodig. Microsoft biedt een gratis programma aan, de "Microsoft Agent Character Editor", dat te downloaden is van de pagina www.microsoft.com/msagent/devdownloads.htm. Het bestand ACE.EXE is maar 1 MB groot. Een korte beschrijving van dit programma is zeer verhelderend om een inzicht te krijgen in de samenstelling van een Agent. In het openingsscherm, zie figuur 3/19.15-33, moet men de eigenschappen van de Agent definiëren. Men voert uiteraard een naam in, de taal waarin de Agent gaat praten, een korte beschrijving en verder wat gegevens zoals of de Agent gebruik moet maken van een tekstballon bij spraak. Bovendien kent het programma hier automatisch een zogenaamde "Identifier" aan de Agent toe.

Dit is een vreselijk lange en ingewikkelde code die van pas komt als de Agent vanuit bepaalde programmeertalen wordt aangeroepen.

19.15 De Agent-technologie van Microsoft

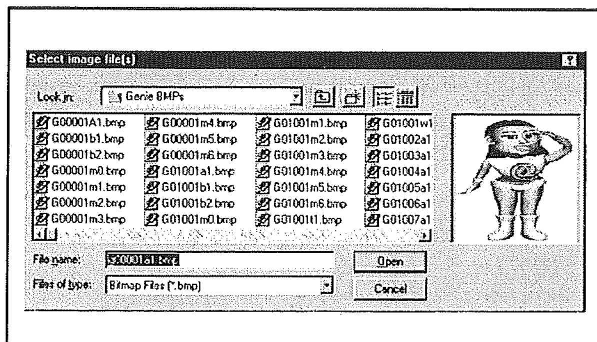


Figuur 3/19.15-33: In het openingsscherm van de "Microsoft Agent Character Editor" moet men de Agent definiëren.

Het vakje "Supports standard animation set" wordt beter niet aangevinkt. Doet men dat wél, dan moet men alle standaard-"States" die Microsoft voorzien heeft voor een Agent aan de animaties koppelen (zie later).
In de tab "Word Balloon" kan men de eigenschappen van de tekstballon instellen, zoals aantal regels tekst, lettersoort en lettergrootte en de kleuren. In de tab "Voice" moet men een van de spraakengines aan de Agent toekennen. Dit is de defaulttaal waarvan de Agent gebruik gaat maken. Dat wil dus zeggen dat applicaties die geen mogelijkheid bieden om een taal en stem in te stellen, hier standaard gebruik van maken.

Tekenen maar!
Vervolgens komt het aan op tekentalent. Elke animatie bestaat uit een aantal BMP-plaatjes van maximaal 128 bij 128 pixels en 256 kleuren. Deze plaatjes worden frames genoemd. Die kan men ontwerpen in een pixelbewerkingsprogramma zoals PaintShop Pro. Het komt er nu op aan een reeks frames te tekenen die telkens iets afwijken van de vorige. Men zou in dit geval van figuur 3/19.15-34 bijvoorbeeld de mond van het meisje kunnen aanpas-

sen, zodat het, als alle frames achter elkaar worden vertoond, lijkt of de Agent spreekt. Elk plaatje wordt onder een eigen naam als BMP opgeslagen. Als alle frames van een beweging zijn getekend, moet deze worden toegekend aan een "Animation".



Figuur 3/19.15-34: Via dit scherm kan men alle getekende frames even snel bekijken en ze in een "Animation" laden.

In het voorbeeld is dus een animatie "Animation 2" geopend (zie "Edit"-menu) en daar zijn reeds twee frames, "Frame 1" en "Frame 2" aan toegekend. Ook het openen van nieuwe frames gaat via het "Edit"-menu. Men kan de "Duration" van elk frame afzonderlijk vastleggen en aan een frame ook een muziekbestand koppelen. Natuurlijk doet men dat meestal alleen bij het eerste frame. Op deze manier ontwerpt men een tiental animaties die elk uit minstens vijf frames bestaan, zie figuur 3/19.15-35. Men kan een animatie afspelen via de optie "Preview" uit het "Edit"-menu. Klaar? Dan moet men de standaard "States" van Microsoft aan de animaties koppelen. Die "States" zijn de standaardhandelingen die een Agent moet kunnen uitvoeren. Hij moet kunnen luisteren ("Listening") of zich kunnen verbergen ("Hiding") en weer tevoorschijn komen ("Showing"). Het zijn deze states met hun standaardbenamingen die worden aangeroepen als men later een bepaalde animatie van een Agent in een programma of op het Internet wil gebruiken. Als men het hokje "Supports standard animation set" niet heeft aangevinkt, hoeft men echter niet alle states te definiëren. Men kiest dan die states uit die voor de toepassing van belang zijn. De Agent is hiermee klaar en via "Build Character" uit het "File"-menu kan men nu de Agent omzetten naar een bestand.

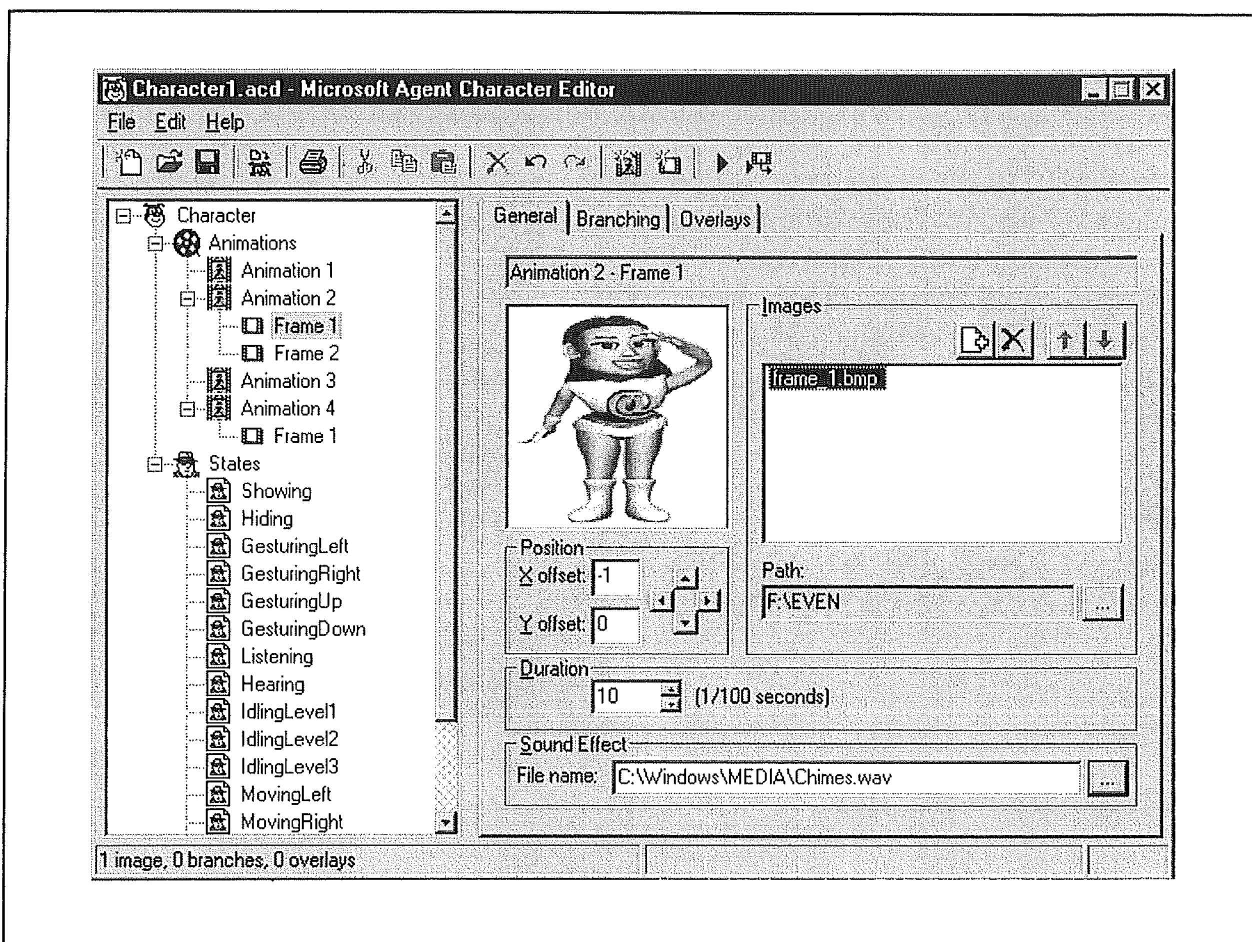
Lipsynchrone spraak

Een unieke eigenschap van de Agents is dat zij in staat zijn de beweging van hun lippen min of meer synchroon te laten verlopen met de woorden die worden uitgesproken. Volgens Microsoft is dat een absolute voorwaarde om de Agents realistisch te laten overkomen. Met de "Microsoft Agent Character Editor" kan men daarvoor zorgen. Men kan namelijk zogenaamde "Mouth Images" maken die over een bepaald frame worden geprojecteerd. Vervolgens kan men een Agent zo programmeren dat voor elk foneem dat wordt uitgesproken een specifiek "Mouth Image" over het basisframe van een animatie wordt gezet.

Volgens Microsoft volstaan zeven van dergelijke "Mouth Images" om redelijk lipsynchrone animaties te ontwerpen, zie figuur 3/19.15-36. Deze hebben een gestandaardiseerde naam waarmee zij kunnen worden aangeroepen:

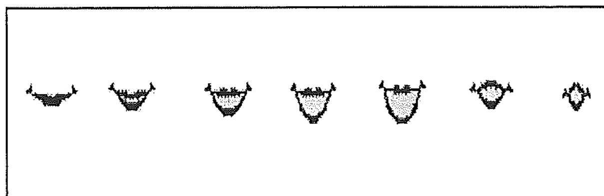
- Closed;
- Wide 1;
- Wide 2;
- Wide 3;
- Wide 4;
- Medium;
- Narrow.

19.15 De Agent-technologie van Microsoft

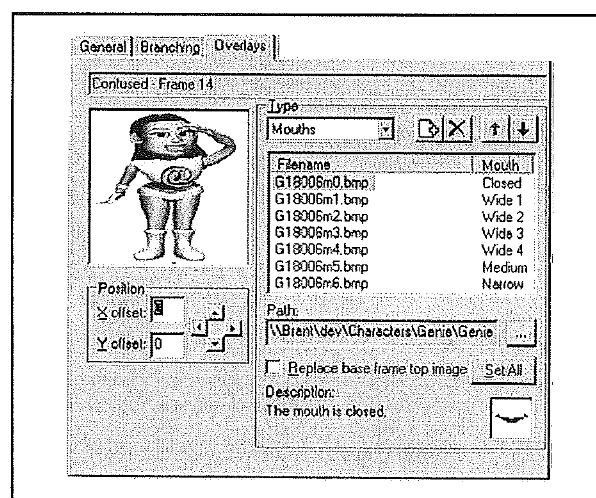


Figuur 3/19.15-35: Monnikenwerk! In dit scherm moet men "Animaties" definiëren en er frames aan toekennen. Nadien moet men de "States" van de animatie koppelen aan de "Animaties".

De "Mouth Images" kunnen aan frames worden toegekend via de tab "Overlays" in het werkscherm van de Microsoft Agent Character Editor, zie figuur 3/19.15-37.



Figuur 3/19.15-36: De zeven "Mouth Images" die Microsoft standaard in een Agent heeft ingebouwd.



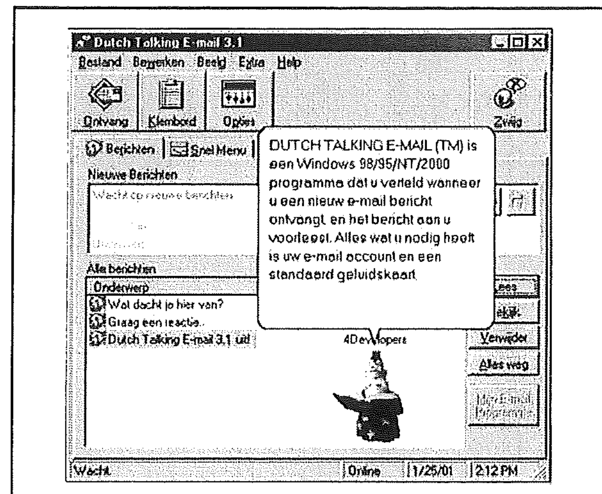
Figuur 3/19.15-37: Via de tab "Overlays" kan men "Mouth Images" over frames monteren.

Een goede tip
Een tip die heel wat werk bespaart. In principe is het samenstellen van een animatie te vergelijken met het maken van een animated GIF. Daar bestaan leuke programma's voor, die bovendien bijna allemaal de mogelijkheid bieden overgangen tussen het ene en het andere plaatje automatisch te laten genereren. Als men hier gebruik van maakt, kan men mooie animaties voor een Agent gemakkelijk ontwerpen. De afzonderlijke GIF's die de frames van de animated GIF vormen, kan men één voor één naar BMP omzetten en men heeft de frames van een animatie klaar!

Andere Agent-toepassingen

Inleiding
Rond het thema "Agent" is een geheel eigen markt ontstaan met honderden aanbieders van Agents en toepassingen die van Agent-technologie gebruikmaken. In deze paragraaf volgt een idee van wat er allemaal is te vinden.

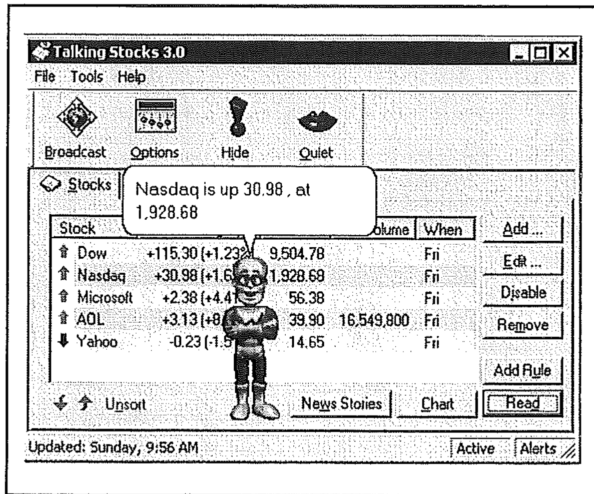
Dutch Talking E-mail
Van de Internetsite van het Amerikaanse bedrijf 4Developers, www.4developers.com, kan men zowaar de Nederlandse applicatie "Dutch Talking E-mail" ophalen, zie figuur 3/19.15-38. Dit is een programma waarmee men e-mailberichten door een Agent kan laten voorlezen. Via een uitgebreid configuratiescherm kan men heel precies instellen wat wél en wat niet uitgesproken moet worden. Met het programma kan men bovendien gesproken reminders opstellen, een pratende klok configureren en zelfs de inhoud van webpagina's laten voorlezen. Via de optie "talking clipboard" kan men teksten die op het klembord staan beluisteren.



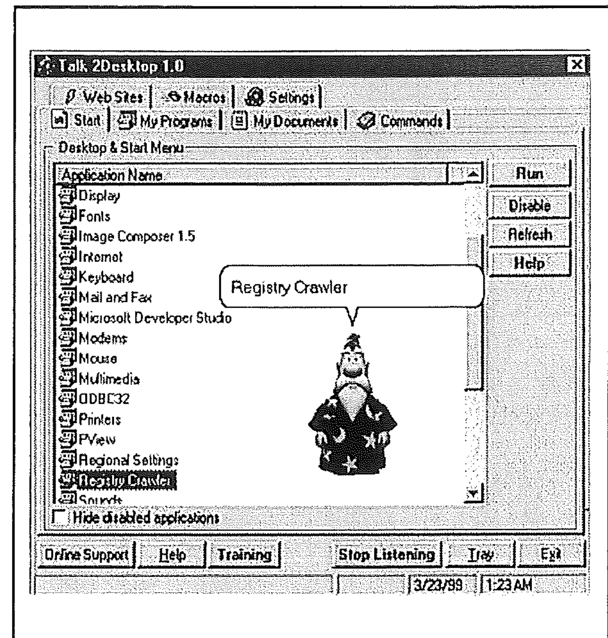
Figuur 3/19.15-38: Met de "Dutch Talking E-mail" kan men ontvangen e-mail berichten laten voorlezen door de favoriete Agent.

Talking Stock
Ook dit is een ontwikkeling van 4Developers en zal zeker de interesse opwekken van iedere aandelen bezitter. Met dit programma, zie figuur 3/19.15-39, legt men contact met bepaalde Internetsites waar de aandelenkoersen up-to-date worden gehouden. Het programma haalt de belangrijkste informatie van deze sites, zet deze in een venster en activeert een Agent die de actuele koersen hardop voorleest. Via een configuratiemenu kan men minimum- en maximumwaarden voor een aandeel instellen. De Agent geeft een gesproken waarschuwing als het aandeel beneden of boven de ingestelde grenzen komt. Men kan het programma zo configureren dat alleen op bepaalde tijden van de dag contact wordt gezocht met de betreffende Internetsites en de aandelenkoersen worden voorgelezen.

19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-39: Met "Talking Stocks" blijft men in real time op de hoogte van de aandelenkoersen.



Figuur 3/19.15-40: "Talk 2Desktop" maakt gebruik van alle aspecten van de Agent-technologie. Het programma levert niet alleen spraakuitvoer via de Agents, maar ondersteunt ook het herkennen van tekst die in de microfoon wordt ingesproken.

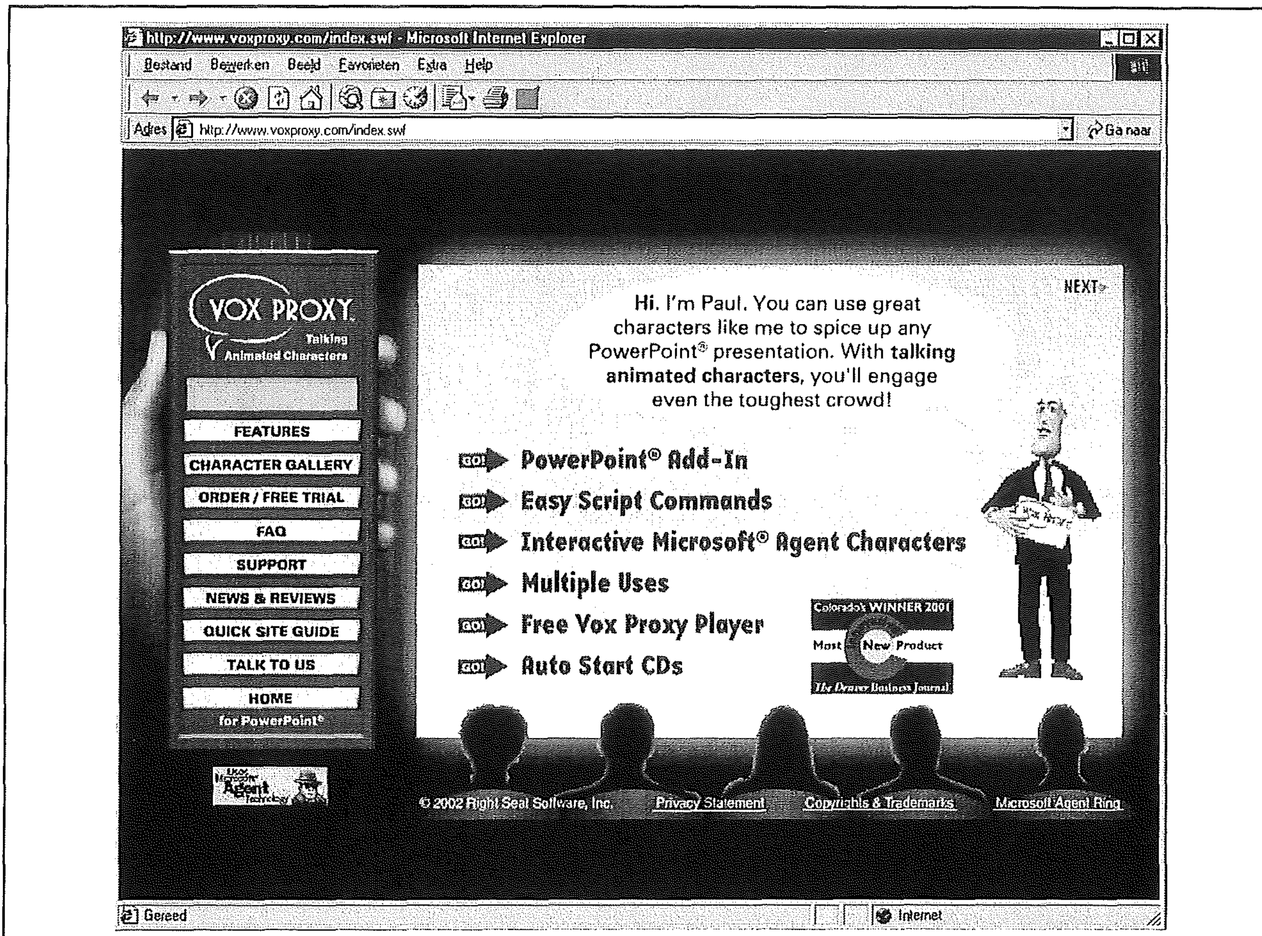
Talk 2Desktop

Ook alweer een ontwikkeling van 4Developers. Deze utility maakt gebruik van de derde, nog niet besproken eigenschap van de Agent-technologie, namelijk de mogelijkheid om tegen de Agent te praten. Met dit programma kan men de computer besturen via de stem. De Agent wordt ingezet als pratende begeleider, die zegt wat men op een bepaald moment in de microfoon moet inspreken om een bepaalde actie te starten. Communicatie in optima forma! Man kan applicaties opstarten, favoriete sites openen, een document bekijken en nog veel meer. Talk 2Desktop ondersteunt tekstmacro's, zie figuur 3/19.15-40, zodat het inspreken van een kort commando volstaat om een ingewikkelde actie uit te voeren. Bovendien kan men op deze manier ook "schrijven". Roep "Aanhef" en het programma weet dat het de bedoeling is dat er een heel lange openingszin van een brief in de tekstverwerker wordt opgenomen.

Vox Proxy

Vox Proxy wordt op de markt gebracht door het gelijknamige bedrijf en men kan het vinden op www.voxproxy.com. Vox Proxy, zie figuur 3/19.15-41, is een ontwikkelomgeving die de Agent-technologie inbedt in PowerPoint-presentaties. Het programma werkt alleen samen met de 2000- en 2002-versies van PowerPoint. In de volledige versie van het programma, die alleen op CD-ROM beschikbaar is, krijgt men 17 Agents, 10 spraakengines en de vrij te distribueren Vox Proxy Player bijgeleverd. In die configuratie kost het programma US$ 249,00. Men kan ook een demoversie downloaden die maar zes Agents bevat, maar dat bestand is wel een sappige 35 MB groot.

19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-41: De site van Vox Proxy, waar men een demoversie van de ontwikkelomgeving kan downloaden waarmee de Agent-technologie in PowerPoint-presentaties wordt geïntegreerd.

De Player, die samen met de PowerPoint-presentatie naar klanten wordt verstuurd, is ook 35 MB groot. Niet geschikt voor Internetdistributie dus, maar die omvang is natuurlijk geen probleem als sprekende PowerPoint-presentaties worden verdeeld via CD-ROM's.

MASH

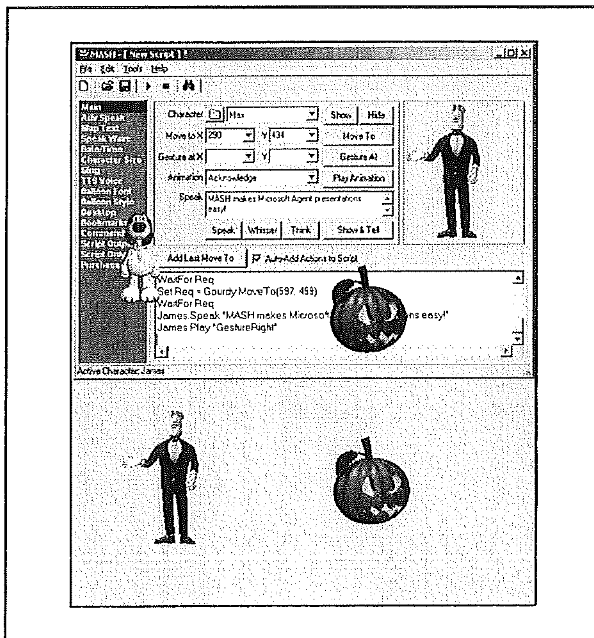
MASH staat voor "Microsoft Agent Scripting Helper". Het is een programma dat door BellCraft is ontwikkeld en helpt bij het integreren van Agents in applicaties zonder dat programmeerkennis wordt verwacht. Men kan ervoor terecht op www.bellcraft.com. MASH ondersteunt VBScript for HTML, JavaScript for HTML, Visual Basic, VDA for Office en het Windows Scripting Host. Met deze applicatie kan men dus ook Agents toevoegen aan Internetpagina's. Het programmeren is werkelijk in een paar uurtjes te leren. Wil men dat Agent James een tekst uitspreekt? Voeg dan de code:

JAMES.SPEAK "Vego is de beste uitgeverij van elektronica boeken"

in en het programma doet de rest.

Wil men dat Agent Robby een bepaalde state uitvoert, oftewel een animatie opstart? Voeg dan de code:

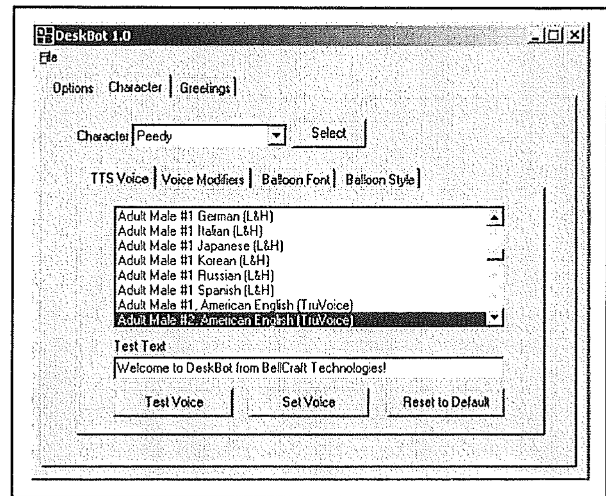
Robby.Play "MovingRight"
in en het programma weet genoeg. Kortom, een programma dat men zonder meer moet downloaden als men serieus met Agents aan de slag wil gaan.



Figuur 3/19.15-42: Met deze "Microsoft Agent Scripting Helper" kan men Agents toevoegen aan HTML-pagina's met behulp van JavaScript of VBScript.

DeskBot

DeskBot is een kleine utility van hetzelfde bedrijf BellCraft dat het toch wel moeizame configureren van alle onderdelen van het Agent-systeem in één venster samenvat. Men kan aan elke Agent een stem toekennen die men vervolgens volledig kan configureren, zie figuur 3/19.15-43. Vervolgens kan men de eigenschappen van de tekstballonnen per Agent instellen. Meer is er over deze handige utility in feite niet te vertellen.



Figuur 3/19.15-43: Met DeskBot vat men alle configuratie-instellingen van de Agent-technologie in één venster samen.

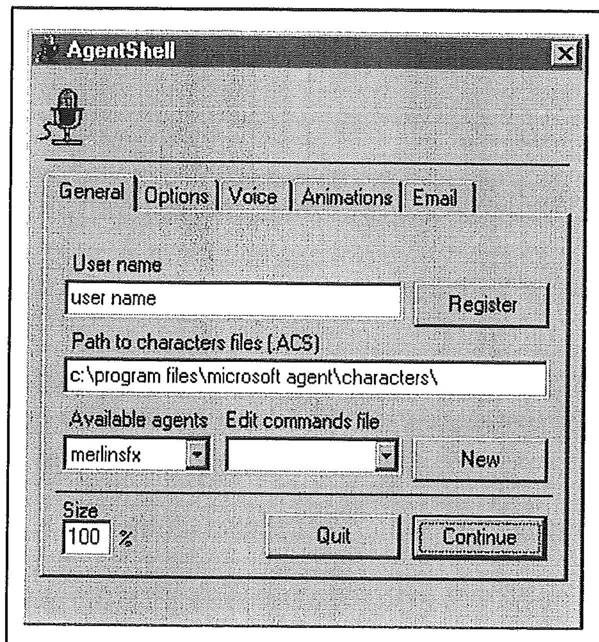
Agent-E-Gram

Een Internetdienst die door BeeBot wordt onderhouden (www.beebot.com) en waarmee men felicitatieberichten waarin Agents zijn opgenomen kan versturen. In een aantal invulschermen vult men in hoe het felicitatiebericht er moet uitzien en nadien wordt het bericht door BeeBot samengesteld en verstuurd.

Agent Shell

Agent Shell van Formedia (www.formedia.com) is een utility die gebruik maakt van de spraakherkenningstechnologie van het Agent-systeem om de PC per spraak te besturen. Men definieert een paar goed herkenbare gesproken commando's en kent daar een bepaalde actie aan toe. Het programma verdwijnt naar de achtergrond maar luistert de uitvoer van de geluidskaart af. Wordt een commando herkent, dan springt het programma tevoorschijn en voort de instructies uit.

19.15 De Agent-technologie van Microsoft



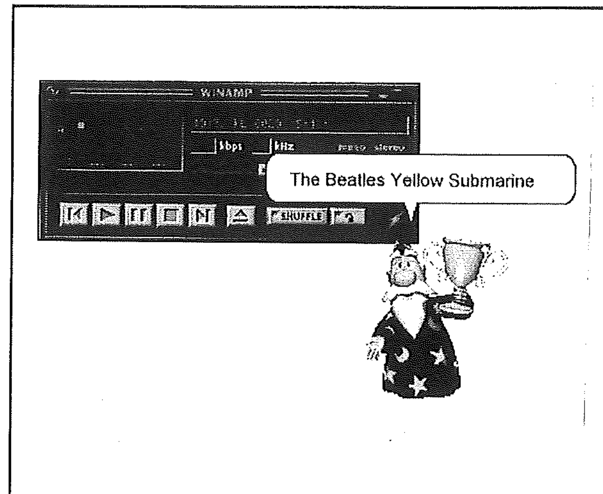
Figuur 3/19.15-44: Met deze Agent Shell kan men de PC gedeeltelijk besturen met gesproken commando's.

Buddy Album Builder

Met deze toepassing van Shelldrake (www.shelldrake.com) kan men een foto-album aanmaken en elke foto voorzien van commentaar, dat uiteraard door een Agent wordt uitgesproken. Men kan het album omzetten naar een EXE-bestand, zodat men het gemakkelijk kan verspreiden. Een nuttige applicatie om bijvoorbeeld een letterlijk en figuurlijk aansprekende catalogus samen te stellen!

Winamp Talking

Met deze utility van MindBeat (www.mindbeat.com) kan men een Agent inzetten als diskjockey bij de bekende MP3-speler WinAmp, zie figuur 3/19.15-45. Iedere keer als WinAmp wil starten met het spelen van een nieuwe nummer uit de playlist, sprint de Agent op het scherm en zegt de naam van het nummer en de informatie die in de MP3-tag is opgenomen. Nadien verdwijnt de Agent weer en start WinAmp met het spelen van het nummer.

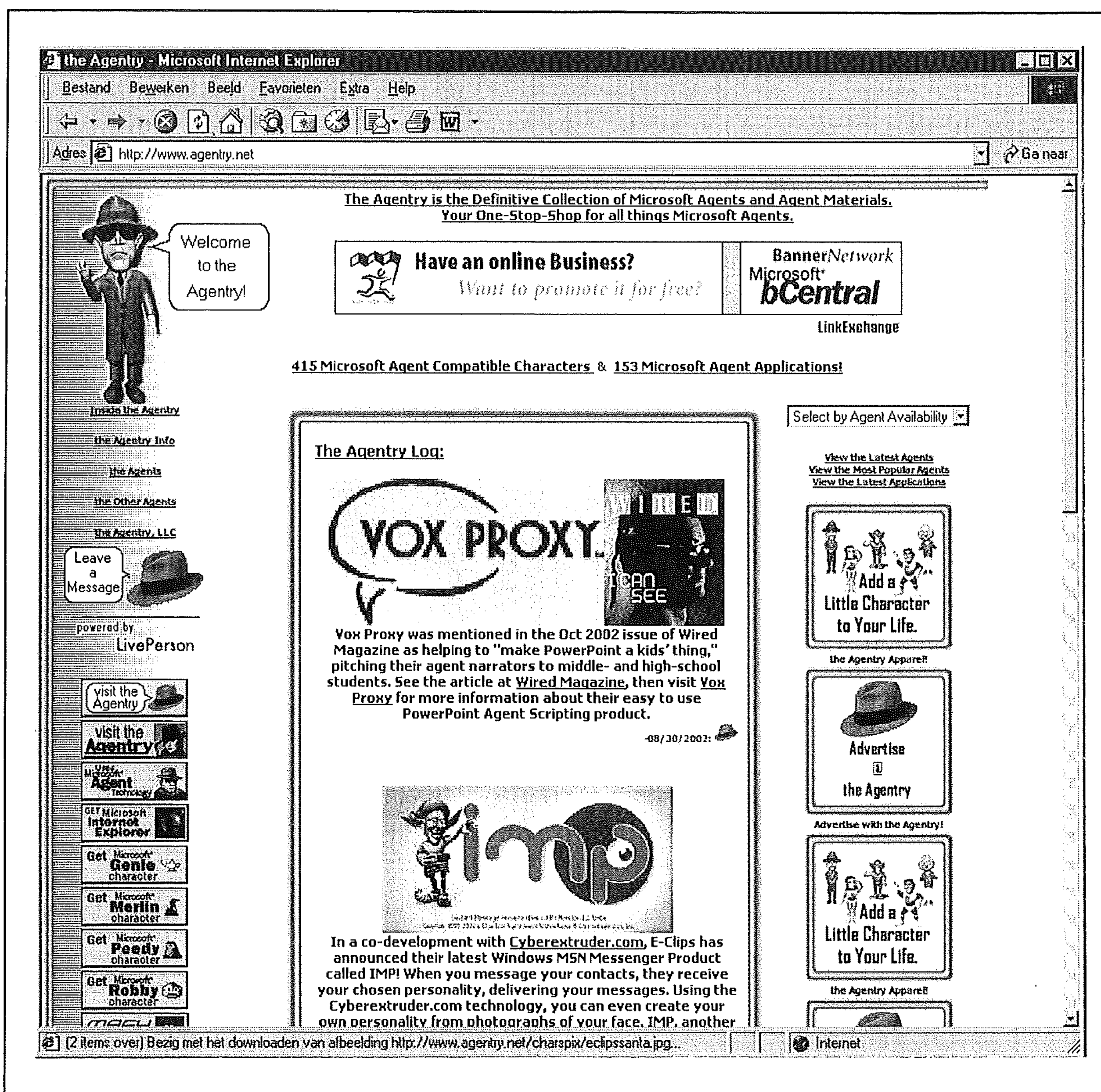


Figuur 3/19.15-45: Met Talking WinAmp kan men de beroemde MP3-player Win-Amp uitrusten met een diskjockey die de titels van de gespeelde nummers afroept.

De Agentry

Tot slot de site www.agentry.net. Hier kan men niet alleen 415 Agents downloaden, maar ook niet minder dan 153 applicaties die van de Agent-technologie gebruik maken. Alle 153 keurig in één lijst met een korte beschrijving en een onmiddellijke verwijzing naar de pagina waarvan men het bestand kan downloaden. Uiteraard zijn er programma's te kust en te keur te vinden die een geschreven tekst omzetten in spraak. Zo is er een utility beschikbaar die uitspreekt wat op het toetsenbord wordt ingetypt. Daarnaast veel e-mail readers, uitbreidingen om Agents toe te voegen aan componenten van Office, om de technologie te integreren in allerlei soorten e-mailprogramma's en links naar programmeeromgevingen waarmee men op een gemakkelijke manier de technologie kan programmeren.

19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-46: De Agentry is een onuitputtelijke bron van rijpe en halfwasse Agent-applicaties.

Agents leren luisteren

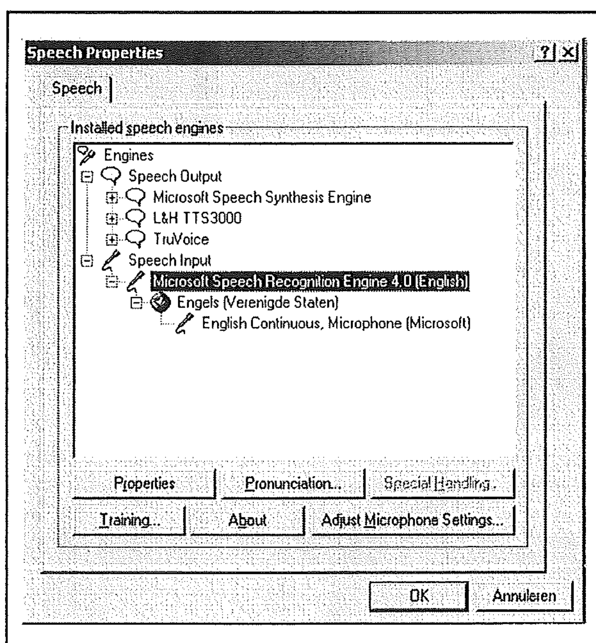
Inleiding

Bij het voorstellen van een paar Agent-applicaties was al gebleken dat Agents ook kunnen luisteren naar commando's die men in een microfoon inspreekt. Dit is inderdaad de derde poot waarop Microsoft zijn Agent-technologie laat rusten. Om met spraakherkenning aan de slag te gaan, moet men echter alweer een programma uit de Microsoft-site downloaden. In dit geval de "Microsoft Speech Recognition Engine".
Men kan deze applicatie downloaden van www.microsoft.com/msagent/downloads.htm.

Het bestand ACTCNC.EXE heeft een omvang van 6.329 kB.

Het "Speech"-paneel wordt uitgebreid

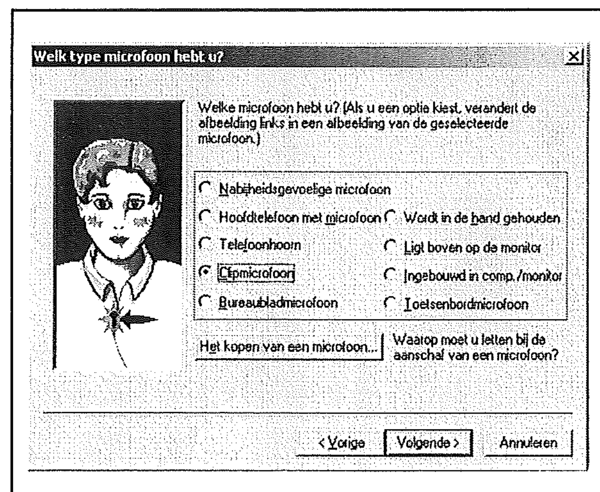
Na de installatie opent men de entry "Speech" van het configuratiepaneel. Daar staat nu een nieuwe engine onder het kopje "Speech Input", zie figuur 3/19.15-47. Bewijs dat het genoemde programma goed is geïnstalleerd. Als men de "Microsoft Speech Recognition Engine" activeert, ziet men dat een aantal knoppen actief wordt. Het zal duidelijk zijn dat men ook deze engine eerst moet configureren.



Figuur 3/19.15-47: De "Microsoft Speech Recognition Engine" is geïnstalleerd en staat ter beschikking in het "Speech"-venster van het configuratiescherm.

Adjust Microphone Setting

Het eerste dat men moet doen is een microfoon aansluiten op de "MIC"-ingang van de geluidskaart en dit apparaatje testen. Daarvoor staat de knop "Adjust Microphone Setting" ter beschikking. In het eerste venster moet men aanklikken wat voor soort microfoon men gebruikt, zie figuur 3/19.15-48. Het meest gebruikte en waarschijnlijk ook het handigste type is de hoofdtelefoon met aangebouwde microfoon. Deze wordt niet alleen aangesloten op de "MIC"-ingang van de geluidskaart maar ook op de luidsprekeruitgangen. Tegenwoordig zijn er echter ook toetsenborden en monitoren in de handel die voorzien zijn van een ingebouwde microfoon.

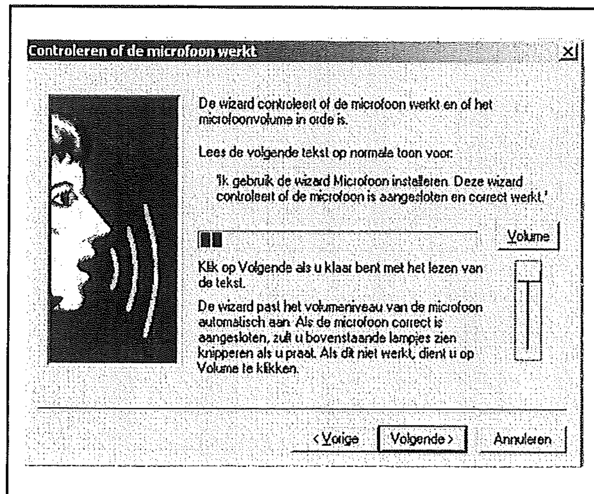


Figuur 3/19.15-48: In dit venster stelt men in welke microfoon wordt gebruikt. Dat is natuurlijk belangrijk, want elke soort heeft een specifieke frequentieweergave en gevoeligheid.

In het volgende venster (figuur 3/19.15-49) wordt verzocht de zin "Ik gebruik de wizard-microfoon installeren. Deze wizard controleert of de microfoon is aangesloten en correct werkt" op een normale manier in de microfoon uit te spreken. Het systeem controleert nu of

19.15 De Agent-technologie van Microsoft

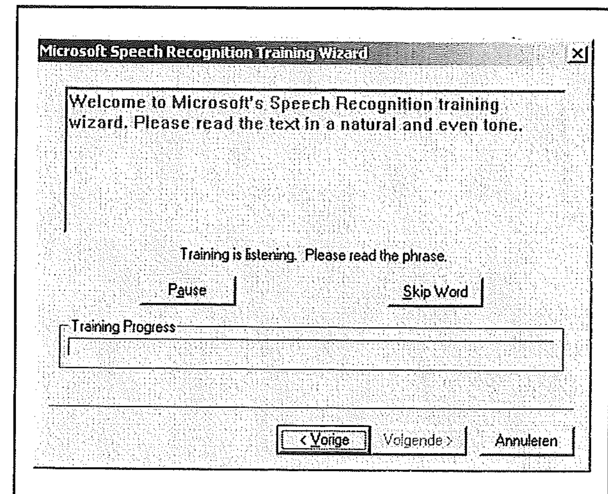
deze zin goed wordt geïnterpreteerd. Is dat niet het geval, dan moet men de procedure herhalen met een andere instelling van het volume.



Figuur 3/19.15-49: Door het uitspreken van een standaardzin kan de spraakherkennings-engine controleren of de gevoeligheid van de microfoon goed is ingesteld.

Training

Vervolgens moet men Engelse uitspraak gaan trainen. Het systeem is namelijk nogal kritisch en men moet toch min of meer op dezelfde manier het Engels uitspreken als het systeem verwacht. Om alvast in de stemming te komen kan men eerst de PC een aantal Engelse voorbeeldzinnen laten uitspreken. Dan weet men alvast wat er wordt verwacht. Nadien moet men een paar eenvoudige Engelse zinnen uitspreken, zie figuur 3/19.15-50. Het systeem geeft aan welke woorden werden herkend. Is de score te laag, dan moet men opnieuw aan de slag tot Microsoft tevreden is met de vorderingen en meldt dat de uitspraak goed genoeg is om de spraakherkennings-engine tevreden te stellen.

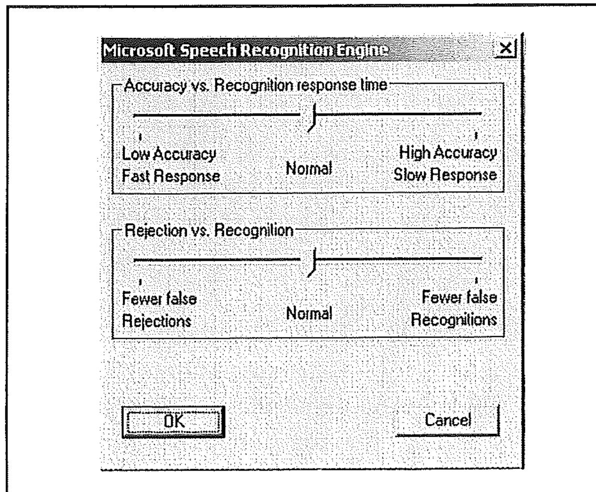


Figuur 3/19.15-50: Oefenen geblazen! In dit en volgende vensters moet men eenvoudige Engelse zinnen uitspreken tot het systeem ze in voldoende mate verstaat.

Properties

Tot slot kan men in het venster "Properties" een bepaalde balans instellen tussen de herkenningstijd en de nauwkeurigheid, zie figuur 3/19.15-51. Het evalueren van spraak kost namelijk, zelfs op een modern snel systeem, toch een bepaalde meetbare tijd. Men kan echter het herkenningsproces versnellen, met de kans dat de uitspraak vaak niet wordt herkend. Met een tweede schuifpotentiometer kan men de verhouding instellen tussen valse herkenningen en valse afwijzingen. Een cryptische omschrijving? Toch is het eenvoudig. Stel dat in een bepaalde spraakherkennings-applicatie zowel aan het woord "open" als aan het woord "over" een bepaalde actie is gekoppeld. De spraakherkennings-engine evalueert de uitspraak en vergelijkt deze met de gegevens die de applicatie in een database heeft opgeslagen. Valse herkenning wil zeggen dat het programma "open" interpreteert terwijl er toch duidelijk "over"

werd ingesproken. Valse afwijzing wil zeggen dat het programma de uitspraak niet herkent en er dus noch de actie "open" noch de actie "over" aan vastkoppelt.



Figuur 3/19.15-51: In dit venster kan men de snelheid van de uitspraakherkenning instellen. Een hoge snelheid gaat meestal gepaard met een hoger foutenpercentage.

Agent-documentatie downloaden

Inleiding

Het zal wel duidelijk zijn geworden dat de Agent-technologie een zeer complex geheel is, waar heel wat grondige studie voor nodig is alvorens er zélf iets mee te bakken valt. Als men met het systeem aan de slag wil, is het ten zeerste aan te raden de uitgebreide documentatie te bestuderen die Microsoft on line ter beschikking stelt. Men kan te rade gaan op de volgende pagina's:

- www.microsoft.com/msagent/otherspeech.htm;
- www.microsoft.com/msagent/characterdata.htm;
- www.microsoft.com/msagent/devdownloads.htm;
- www.microsoft.com/msagent/documentation.htm;
- www.microsoft.com/msagent/devnotes.htm.

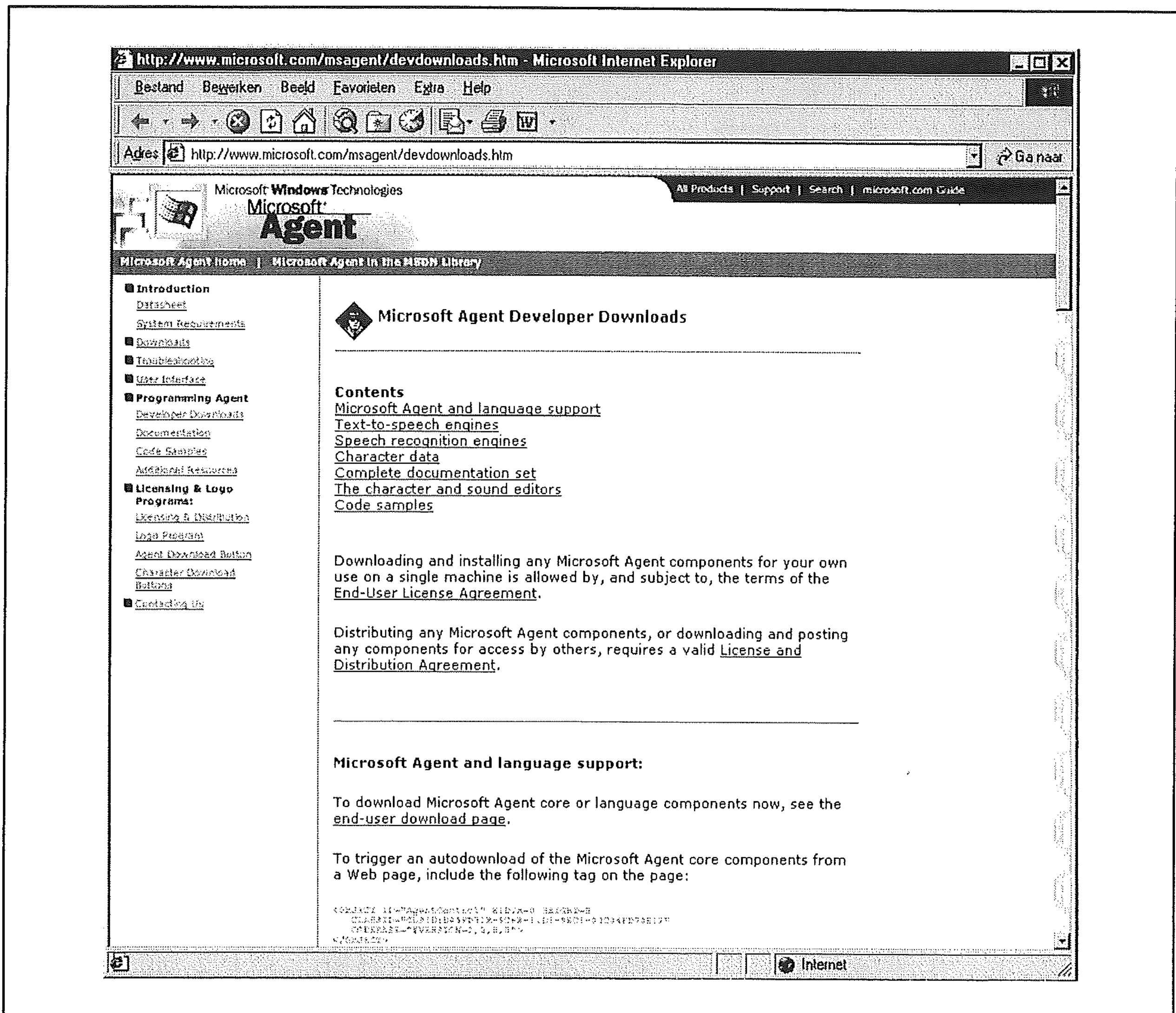
Documentatie in Word-formaat

Van www.microsoft.com/msagent/devdownloads.htm kan men de volledige documentatie in Word-formaat downloaden. In dit zelfuitpakkende EXE-bestand zitten niet minder dan 25 tekstbestanden verborgen met een gezamenlijke omvang van 6,7 MB, goed voor honderden pagina's interessante lectuur. Alle aspecten van het gebruik en het toepassen van Agents komen aan de orde, tot en met een zeer uitgebreide handleiding: "Programming the Microsoft Agent".
In dit document leert men alles over het programmeren van de Agents in HTML-pagina's.

Codesamples

Op dezelfde pagina staan links naar codesamples in HTML, Visual Basic, Visual C++ en Microsoft Office Environment die men kan toepassen en bestuderen.

19.15 De Agent-technologie van Microsoft



Figuur 3/19.15-52: Van deze Agent-pagina op de Microsoft-site kan men uitgebreide documentatie over de technologie downloaden.

19.15 De Agent-technologie van Microsoft